A METHOD FOR TONIC SELECTION FOR INDIAN MUSIC SINGERS

A Thesis Submitted

in Partial Fulfillment of the Requirements

for the Degree of

Master of Technology

by

Chandrakant J. Gaikwad

to the

DEPARTMENT OF ELECTRICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY,KANPUR

July 2005

CERTIFICATE

It is certified that the work contained in the thesis entitled "*A METHOD FOR TONIC SELECTION FOR INDIAN MUSIC SINGERS*" by *Chandrakant J. Gaikwad* has been carried out under my supervision and that this work has not been submitted elsewhere for a degree.

Pradip Sircar

(P. Sircar)

July 2005

Professor,

Department of Electrical Engineering,

Indian Institute of Technology,

Kanpur-208016.

Abstract

A method for tonic(sa) selection is discussed. Singers were asked to select three tonics of their choice. Then they were asked to sing aaroh and avaroh with the selected tonic in akaar(aalap) i.e. without pronouncing any syllables, only using the sustained /aa/ vowel sound. Notes in the aaroh and avaroh with these three different tonics were analyzed for their timbre(quality).

In this experiment several audio clips drawn from commercial recordings of professiona singers (e.g. Lata Mangeshkar, Md. Rafi, etc.)are used for illustration of criterion, used for the tonic selection. Spectral domain techniques and autocorrelation based pitch detection algorithm is used to analyze the musical notes.

A tristimulus method suggested by Pollard and Jansson(1982) for the specification of musical timbre is used to represent the timbre of the notes sung by the singer. Timbre(quality) of the notes is compared using tristimulus diagrams. Tristimulus diagrams are drawn for all the notes in the aaroh and avaroh. Position of these notes in the tristimulus diagram determines their timbre(quality)

Using classical timbre theory and analysis of Indian music singers voices, voice range and tonic is determined.

Dedicated to my father

Late Jagannath Laxman Gaikwad

Acknowledgements

I am grateful to Prof. P. Sircar for giving me an opportunity to work on the problem of Tonic selection for Indian music Singers and for the support and motivation throughout the course of the thesis.

I wish to thank the various singers and artists who took the time and effort to meet and discuss music with me during the course of thesis, including Adish Vartak, Rajendra Singh and Kedar Jathar who were all kind enough to record some audio samples for me. I also thank ethnomusicologist Dr SAK Durga , for finding time in her busy schedules to meet with us for discussions.

In addition I also thank the faculty and the Department of Electrical Engineering, IIT Kanpur in general for the help and resources throughout the course of my master's program.

To my colleagues Ritesh, Vinay, Vijay, Dilip, Hari, Hemant, Jawahar, Mayank, MKS, Pulkit, Raghu, A T Rajesh, Ravindra Lokhande, Suresh, Kiran thanks for your wonderful company and help during my stay at IIT Kanpur.

Most importantly, I thank my elder brothers and my sisters for their unconditional love and support.

Contents

List of Figures

Chapter 1

Motivation

Indian music is based on the basic shadja, also called as tonic or Sa. A raga can be identified only if the Sa note pitch(tonic) is identified as the other notes in the scale are related to the basic shadja. The quality of the voice depends on the choice of the basic Sa as Indian music singing requires a voice which sounds pleasant and to be heard with power in all the three registers or octaves. Voice is not classified in Indian music singers as we find in western music as tenor, baritone, bass, soprano. But Indian music singers have freedom in choosing their basic tonic note (Sa).

The tonic is not chosen by any logic or scientific method and it is taken by the singers at random. In the sense, if the lower pancham(Pa) is not heard, the tonic note is raised and if the higher pancham note (Pa) is not reached, the singer lowers his tonic. Thus the singer gropes to find his right tonic and finds difficult to sing either the lower octave or higher octave. Besides this problem his or her voice quality suffers and lacks resonance and results in a bad voice.

Though the riyaz is done regularly, the sadhana becomes an incorrect practice and makes the voice quality poor so the motivation behind this work is to propose a

scientific method for tonic(Sa) selection.

Chapter 2

Fundamentals of Indian Music

2.1 Concept of Swara, Aaroh/Avaroh, Octave & Scale

Any melody is a progression of an up and down flow of sound along the stream of time. It is obvious that this melody though simple, is the beginning of all later complex ragas. The breakdown of the 'up and down' movement yields the notes(swaras) and a series of swaras arranged in a certain order within certain limits is a scale. In the following subsections these concepts are explained in detail.

2.1.1 Primary Notes(Swaras)

Indian music is based on seven primary Swars(notes). Of these, two, the shadja(the tonic) and the panchma(the fifth) are fixed notes, in the sense that they have no variations (as flat and sharp) as the rest of the five notes have. Each of the four notes Ri, Ga, Dha and Ni has a komal variation which is lower in pitch than the origional note.

The remaining note Ma has however, a tiwra(sharp) variety. This makes a total of

twelve notes in an octave (Saptaka).

The most important note is the fundamental note Shadja or the Tonic. No musical composition is conceived without a tonic as no structure is contemplated without a base. The tonic determines the relative pitch of all notes in the octave. This explains why a vocal or instrumental piece of music is always played to the accompaniment of a drone instrument. Notice, however, that no particular note of definite pitch is specified as the tonic. One is free to choose, according to one's voice-register, any suitable note as one's tonic.

Table I below shows the notes, their names in Hindustani music, varieties, order and notation used in this thesis:

Sr.No.	Hindustani Name	Abbreviation	Notation Used
1	Shadja	Sa	S
2	Komal Rishabh	ri	r
3	Shuddha Rishabh	Ri	R
4	Komal Gandhar	ga	g
5	Shuddha Gandhar	Ga	G
6	Shuddha Madhyam	Ma	M
7	Teevra Madhyam	ma	m
8	Pancham	Pa	P
9	Komal Dhaivat	dha	d
10	Shuddha Dhaivat	Dha	D
11	Komal Nishad	ni	n
12	Shuddha Nishad	Ni	N

Table 1.1: Notes and Their Names

2.1.2 Aroh, Avaroh & Octaves

As we ascend from Sa to Ni the pitch becomes higher and higher and this ascending sucession of notes is called aaroh. In the reverse order, i.e., decending sucession of notes is called avaroh. Each note has some fixed relation to the basic note or tonic, Sa. When we ascend from Sa to Ni, the next suceeding note after Ni is again Sa. Similarly all further notes repeat themselves in succession. So also, when we descend from the basic note Sa, the next note downwards is Ni and similarly all other notes are repeated in a descending order. These octaves or Saptakas are styled as

1. Middle, the Madhya,
2. Higher, the Tara,
3. Lower, the Mandra.

Now the pitch of any note in tara saptaka is exactly the double of its identical note, i.e., its octave in the madhya saptaka and the pitch of any note in mandra saptaka is half of its octave in the madhya saptaka. Though the pitch differs, the sound of one note in one saptaka and its octave in the other saptaka is identical.

The range of an octave is also known as sthayi. If one starts on any note in a natural pitch, it is called the beginning of one's middle octave, madhya sthayi. Lower to this is mandra sthayi. As this goes up the scale and completes the saptak and goes to the next, one is said to enter into the tara sthayi (upper octave). And, obviously, there can be progressively downward and upward sthayi-s.

2.1.3 Notation Used

Notes shown in the table I are the notes in the middle octave or madhya saptak. Notes in the lower octave or mandra saptak will be preceded by ‘ & notes in the higher octave or tar saptak will be succeded by ’.

e.g. ‘N represents Nishad or Ni of lower octave and S’ represents Sa of higher octave.

2.2 Western Music, Indian Music and Keyboard

Even though Indian musical systems are very different from the traditional Western music system, we can still get a lot of insight into Indian music by studing equally tempered, twelve keys per octave methodology.

2.2.1 Keyboard and ‘Equally Tempered’ Arrangement

The audible frequency range is divided into ‘octaves’. An octave is a frequency range from a frequency f_1 to f_2 such that f_2 is twice that of f_1 in terms of cycles or hertz. We can choose any number to be our f_1 (and f_2 of course is 2 times f_1)-we can define an octave from 240 Hz to 480 Hz or equally well another one, say from 120 Hz to 240 Hz.

A piano or a keyboard is a typical Western musical instrument([23]). All we see is a bunch of keys, some in black and some in white. However, upon a closer look, we see that there is a periodicity. As we go from the left of the keyboard to the right the keys produce higher and higher frequencies. In fact, the key frequencies are arranged in such a manner that they are in a geometric series. That is, the frequency between any key and the key immediately to its left (irrespective of the color of the key) is a constant, the constant being equal to the twelfth root of two or 1.059. For example,

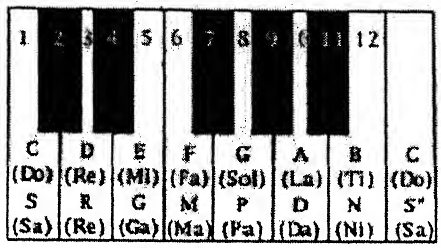


Figure 2.1: Keyboard

typically, there is a white key in the keyboard set to 240 Hz. Then the adjacent key on the right, a black one as a matter of fact, is set to 240 X 1.059 = 254 Hertz.

By the specific choice of this ratio (twelfth root of two) we see that by the time we reached the thirteenth key, we have doubled our frequency and thus spanned a whole octave. In fact, if we look at the keyboard we see that the key pattern repeats every twelve keys. If we chose the white key at 240 Hz, then the thirteenth key will be at 480 Hz and our octave ranged from 240 to 480 Hz. Equally well, we could have started counting from the black key at 254 Hz and twelve keys later we would have still spanned an octave, except that this time our octave ranged from 254 to 508 Hz.

This division of the octave into twelve 'tones' which have specific ratio between adjacent keys (the ratio equalling 1.059) is peculiar to Western music. This geometric arrangement of frequencies of the keys in an octave is called an 'equally tempered' arrangement. And besides the keyboard, most Western musical instruments are also tuned to such an arrangement.

Even though there is a degree of freedom about what we want to be the range of an octave (whether it is from 240 to 480 Hz or 254 to 508 Hz etc.) the Western music defines a standard octave called the 'Middle C octave' (also called the Middle C

scale,etc) starting from the white key set to 240 Hz. The entire octave (the twelve key pattern) is shown in Table II. On keyboard, this octave is located near the middle.

The upper octave, starting from 480 Hz is the Upper C octave and the lower octave starting at 120 Hz is the Lower C octave etc.

From Table II given below, we notice that the keys in the octave have labels for identification. Of the white keys - there are seven of them in an octave - the first one is called C (and hence the name 'Middle C' octave) and then we progress alphabetically to G and then back to A and B, after which, the present octave ends and the C key of the next octave begins. The same labeling system is repeated for the keys in the other octaves as well. The five black keys have ambiguous labels, because each one of them has two labels. The first black key, for example, is called 'C sharp' (C #) or 'D flat' (Db) - it is obvious that 'sharpening' essentially is a technical term for being 'one key higher' and similarly 'flattening' is one key lower in frequency than the white key in the prefix. The labels, frequencies etc of all the twelve keys in the Middle C octave are provided in Table II.

Key #	Key color	Frequency (Hz)	Notation Used
1	White 1	240	C
2	Black 1	254	C # (D b)
3	White 2	269	D
4	Black 2	285	D # (E b)
5	White 3	302	E
6	White 4	320	F
7	Black 3	338.5	F # (G b)
8	White 5	358.5	G
9	Black 4	380	G # (A b)
10	White 6	402	A
11	Black 5	426	A # (B b)
12	White 7	451	B

Table II: Arrangement of keys in a keyboard

By definition, each key is supposed to be a 'semitone' or 'half tone' apart from its adjacent key. Thus, keys which are second nearest neighbors are considered a 'whole tone' apart.

For example, the first white key ('C' key) and the first black key ('C sharp') are a 'semitone' apart, whereas the first white key ('C key') and the second white key ('D key') are a full tone (whole tone) apart. And the 'C sharp' and 'D' keys are a semitone apart, as well.

The traditional Indian music system is based on a 22 key per octave system. and scales used are different from 'equally tempered' arrangement. They are called 'Just tempered scales'.

2.2.2 Western Versus Indian Classical Music System

We noted that in Indian music it is not enough to produce just twelve 'tones' in an octave. One ought to produce even the intermediate frequencies. These intermediate frequencies, which do not have any keys to produce them, are called 'microtones'. The Indian word for the 'microtone' is 'gamak'. Microtones add variety to the Indian classical music - an extra dimension. From movie songs to folk music to classical music, the very heart of Indian music is this 'continuous flow' or 'gliding through a continuum of frequencies' or gamaka or microtonal excursions. Thus it is often said that Indian music is 'melody-based'. Since microtones are so important in Karnatic and Hindustani music and very few instruments can produce all the frequencies in an octave, the best enunciation of Indian classical music is in vocal singing. Western music is 'harmony-based', which brings out yet another difference between the two systems. 'Harmony' is produced when several instruments play different melodies or pieces simultaneously like in an orchestra. Harmony is also produced when more than one tone is produced at the same time. In the Western Music, 'harmony' is an important element. Orchestration and 'harmony' are absent in Indian classical music. Indian classical music, does not use what are called chords, or pressing more than one key simultaneously. Chords are a major aspect of Western music and producing harmony via chords is a natural consequence of the equally tempered (geometric series) arrangement of the keys. If keys were arranged in a Just tempered sequence, pressing more than one key at a given time might produce an unpleasant sound pattern resulting in what is called 'Besur' (in Hindustani music) or 'Abaswaram' (in Karnatic music).

Advantage of Equal temperment of pianos and keyboards is that it makes it easier

to 'tune' them, (they go out of tune every once in a while and need to be tuned periodically) since each key is harmonically related to the other keys. In case of Just tempered arrangement, since the key ratio between adjacent keys is not a constant, most keys will have to be tuned individually.

Also, the Western scales are standardized. The middle C octave ranges from 240 to 480 Hz. In Indian music, we have the freedom to choose the frequency range of the octave from anywhere to twice anywhere. We can start at 230 Hz, if we wish.

Just to summarize([23]), the essential differences between Indian classical music system and the Western music are:

(a) the Western keyboard is 'equally tempered' whereas the Indian keyboard ideally should be 'just tempered'.

(b) Only twelve keys per octave are used in the West, whereas to play Indian music one needs to produce several intermediate microtones, not represented by a conventional keyboard - This is the most major difference.

(c) Harmony, chords, polyphony etc are absent in Indian classical music.

(d) In Indian music, there is no need to standardize an octave to begin at 240 Hz.

In Indian music system, we do not use alphabets to label keys. Instead, we use short, syllables which go - Sa ri ga ma pa dha ni. These seven syllables are actually mnemonics to represent the 'notes' or 'swaras' in Indian music. They are referred to as the 'Saptha Swaras' or 'Seven Swaras'.

This notation (and this set of seven 'notes') is also called the 'solfege notation' in the west which goes do, re, me, soh etc. Basically, the solfege notation is a 'singable' set of syllables which helps us describe a musical melody. Many good Indian musicians

have voices spanning the entire three octaves, although most Indian compositions use up just the complete madhya stayi scale and the top half of the mandra stayi (only half an octave below) and the bottom half of the tara stayi (just half an octave above the madhya stayi).

We also see that the twelve keys of the octave divide into two halves. The four keys which are designated as ri and ga are called the 'bottom tetrachord' (in Indian terminology, 'poorvaangam') and similarly the four keys corresponding to dha and ni are called the 'upper tetrachord' or 'uttaraangam'.

2.3 Comparison of Scales

Below is a table comparing various scales([14]), including Equal Temperament, Pythagorean, Natural Tuning, and commonly used Bhatkhande's "Indian" scale. The discussion is carried out using the key of C as example.

Equal Temperament:

Divides the octave into 12 equal semitones, each spaced at a ratio of 1.05946 (being the 12th root of an octave, 2).

Pythagorean:

Builds ratios on the pure perfect fifth (3:2), scaling back into the appropriate octave by dividing by 2, 4 or so on. So D is 9/8, being 3/2 * 3/2, lowered back in to this octave by dividing by 2. Similarly, E is 81/64.

Natural:

This scale is based on simple ratios and includes the Classic "Just" Diatonic scale. Whereas Pythagorean only allowed 3 as the highest prime, this one allowed up to 17.

Variations are achieved by changing the upper limit.

Shrinivas and Bhatkande:

These are the two Indian musicologists who made great inroads into systematizing Indian music. Their definitions of intervals were expressed as lengths on a string, perhaps in reference to frets on, say, the veena or sitar. In Indian music Pandit Bhatkhande's scale is commonly used. The ratios for Pythagorean and Natural scales are included for comparison.

Indian Scale	Eq.Temp	Pythagorean	Natural	Bhatkhande	C Scale
Sa	1	1	1	1	C
re	1.06	256/243	16/15	256/243	Db
RE	1.12	9/8	9/8	9/8	D
ga	1.19	32/27	6/5	6/5	Eb
GA	1.26	81/64	5/4	5/4	E
MA	1.33	4/3	4/3	4/3	F
ma	1.41	729/512	17/12	45/32	F#
PA	1.5	3/2	3/2	3/2	G
dha	1.59	128/81	8/5	50/31	Ab
DHA	1.68	27/16	5/3	27/16	A
ni	1.78	16/9	9/5	9/5	Bb
NI	1.89	243/128	15/8	15/8	B
SA'	2	2/1	2/1	2/1	C

Table III: Comparison Of Various Scales

Collectively the notes SA re RE ga GA MA ma PA dha DHA ni NI SA are known as the sargam which is somewhat analogous to the Western solfege: Doh Re Mi Fa Sol La Ti Doh – but not quite. The Western solfege scale usually refers to the tempered scale (c d e f g a b – as on the piano) and the Eastern scale usually refers to the

"Natural" or "Harmonic" scale. The notes in the Western scale are evenly spaced, the ones in the Eastern scale follow the natural divisions of vibrational frequencies.

2.4 Qualities and Defects of Notes in Indian Music

Musicologist B. Joshi([8]) described some qualities and defects of notes in his book *'Understanding Indian Music'*. He says to be musically fit, the note must not only be melodious, but it must possess many more qualities and must be free from a number of defects. Some of these qualities and defects are as follows:

Note must remain steady i.e. must not fluctuate, flicker or crack. Its intensity also must remain constant. If it is otherwise, that necessarily mars its sweetness and beauty. It must also be a prolonged note. Long drawn notes produce a deeper and more sustained effect than short notes. The long notes sung in low rhythm impress deeply and have a better staying effect.

Not only must a note be steady and sustained but continuity of voice must be maintained while improvising, i.e., while rendering various notes the breath must be sustained, without any break, for a pretty long time as far as practicable.

Another important quality to be achieved is the intensity or volume of the voice. By volume, not only is the audibility of the note increased, but its effect on the ear is also deepened, as the impact of an intensive voice is bound to be greater. These qualities of the note are quite essential for scientific music which has to create a deeper and serious effect on the listener.

The sound produced must be clear, free, and full. It should not be nasal , throaty or husky, nor should it be produced by jerks. At the same time it should not be harsh

but soft.

Chapter 3

Past & Recent Research and Feature Extraction

3.1 Dimensions to Sound Perception

Generally, a musical sound can be described by four factors:

1. Pitch,
2. Intensity(Loudness),
3. Duration, and
4. Timbre.

The first three terms are believed to be one-dimensional, and are better understood primarily due to the existence of their physical correlates. That is, pitch is measured in terms of fundamental frequency, loudness explained through intensity, duration determined by the lifetime of a tone or musical phrase.i.e. These factors can be described as:

Intensity: It is same as loudness and is related to amplitude of the sound wave.

Duration: It is simply the time during which the specific frequency or tone lasts.

Pitch: Pitch is the perception of the frequency of a note.

Timbre: It is a signature of the source of the sound. When voice or instrument produce sound it produces a spectrum consisting of several overtones along with fundamental frequency. This is referred to as timbre or tone color. This constitutes the quality of that sound. Timbre is multidimensional in nature.

3.2 Qualities and Defects of Notes in Indian Music

In the last chapter, we have seen that, the note should possess many qualities and should be free from number of defects. Some of these qualities and defects are as follows:

1. Note must remain steady i.e. must not fluctuate, flicker or crack.
2. Its intensity also must remain constant.
3. It must also be prolonged note.
4. The sound produced must be clear, free, and full.
5. It should not be nasal , throaty or husky, nor should it be produced by jerks.
6. At the same time it should not be harsh but soft.

In the following subsections past and recent research in timbre is reviewed to correlate these qualities and defects of notes with physical parameters (e.g. harmonics present and their relative strengths, duration,intensity, etc.)

3.3 Research in Timbre: Past and Recent

It was Helmholtz([1]) who first attempted a systematic explanation of musical quality in terms of harmonics. He insisted that differences in quality were all capable of explanation in terms of the particular selection of partial tones associated with any note and their relative intensities. Helmholtz stressed the importance of the musical tone, which continues uniformly, i.e. the steady state part of a tone, disregarding peculiarities of beginning and ending thereby neglecting some of the temporal aspects of musical tones. The concept behind his thoughts became to be known as the classical theory and has without a doubt contributed greatly to the research in timbre.

3.3.1 Helmholtz's Conclusions

1. Single simple tones have a very soft, pleasant sound, free from all roughness, but wanting in power, and dull at low pitches.

2. Musical notes which are accompanied by a moderately loud series of the lower partial tones up to about the sixth are more harmonious and musical.

3. Compared with the single simple tones above are rich and splendid, while they are at the same time sweet and soft if the higher partials are absent.

4. If only the odd numbered harmonics were present, the quality of the tone was hollow, and when a large number of such upper harmonics were present, it was nasal. Dominance of odd harmonics is also a feature of square waves which can be represented as the sum of odd harmonics with a decrease in amplitude for each harmonic.

5. When the fundamental tone predominates, the quality of tone is rich, but when the fundamental is weak, the quality is poor.

6. When partials above the sixth are prominent, the quality is cutting and even rough.

7. The musical tones of the same quality would always exhibit the same combination of partials.

8. Nearest to the musical tones without any upper partials are those with secondary tones which are inharmonic to the prime.

3.3.2 Role of Deviation from Exact Locations of Harmonics

Charles Culver([11]) asserts:

When the frequency of one or more upper partials is not exact multiples of the fundamental, if the discrepancy is not more than a few cycles the quality of the tone will not be seriously impaired. If however, the departure from being an exact multiple is appreciable such an overtone constitutes an inharmonic partial and the resultant complex tone becomes rough and hence unpleasant. Inharmonic partials, in general have relatively high frequencies.

3.3.3 Harmonics and Specific Qualities

Jeans(Science & Music, p.86) correlates harmonics with the specific qualities that they represents. According to him:

1. The second partial adds clearness and brilliance.

2. The third partial again adds brilliance, but also contributes a certain hollow, throaty, or nasal quality.

3. The fourth adds yet more brilliance, and even shrillness.

4. The fifth adds a rich somewhat horn-like quality to tone.

6. The sixth adds a delicate shrillness of nasal quality.

7. These six partials are all parts of the common chord of the fundamental, but this is not true of the seventh, ninth, eleventh, and higher odd numbered partials, these add dissonance and introduce a real roughness or harshness.

3.3.4 Consonance, Dissonance and Roughness

When two or more tones, evoked simultaneously produce a rough auditory sensation, it is said that the sounds involved are dissonant([1]), when auditory roughness does not obtain, the sounds are classified as being consonant. Dissonance implies harshness. When notes C & E simultaneously sounded on the piano for instance, produce a smooth musical effect, while the sounding of C & D has quite the opposite effect. Answer to the above question was given by Helmholtz. It was his judgment that dissonance is due to the disagreeable sensation produced by rapid beatings in auditory peripheral channels. Helmholtz explained the perception of musical dissonance in terms of two simultaneously sounding musical tones. These beats could result in intermittent neural activity. According to Helmholtz, consonant intervals were pleasant because very few beats were produced in auditory channels.

3.3.5 Effect of Strong Odd/Even Harmonics

If odd numbered harmonics are weak, the pitch f_o, which is perceived on the basis of the lower even numbered components is too high. The auditory system fails to perceive pitch sensation of $f_o/2$ because odd numbered components are weak and masked by adjacent harmonics. The low pitch notes can be perceived as an octave higher if the even numbered harmonics are fairly weak([2]).

3.3.6 Effect of Higher Harmonics

On the distribution of the harmonics, it has been suggested that no harmonics higher than the 5th to 7th, regardless of the fundamental frequency, are resolved individually. Studies have shown that the upper harmonics rather than being perceived independently are heard as a group (Howard, Angus 2001). Further support for this phenomena is made by Hartman who, according to Puterbaugh (Puterbaugh 1999), suggests that for a signal with fundamental frequency below 400 Hz, only the first 10 harmonics play an individual role: harmonics greater than 10 affect the timbre en masse.

3.4 Feature Extraction

As discussed in previous sections, to be musically fit, the note should possess many qualities and should be free from a number of defects.

Some of these qualities and defects are as follows:

1. Note must remain steady i.e. must not fluctuate, flicker or crack.
2. Its intensity also must remain constant.
3. It must also be prolonged note.
4. The sound produced must be clear, free, and full.
5. It should not be nasal , throaty or husky, nor should it be produced by jerks.

We will analyze notes for these parameters and from these parameters we can get range of the singer and using all these information we can find out tonic or sa of the singer. Detailed method is explained in chapter 4.

From the pitch plot(function of time) we can get information about steadiness and

the duration of the notes. From the previous sections it is clear that the remaining parameters (attributes of timbre or quality of notes) depends on

1. Fundamental frequency,
2. Number of harmonics, relative strengths of harmonics
3. Inharmonic Partials and
4. Spectrum change over time.

These parameters are obtained using harmonic analysis of notes sung by the singers. Finally using 'Tristimulus method' we can obtain spectrum change over time.

So in this thesis basic signal processing techniques used are:

1. Pitch Determination Algorithm (Autocorrelation Method)
2. Harmonic Analysis Using Discrete Fourier Transform
3. Tristimulus Method For Singing Voice Timbre Representation

3.5 Pitch Determination

Pitch, i.e., fundamental frequency(or rate of vocal fold vibration) F_0, as well as fundamental period T_0, has a key position in the music and speech signals. The ear is by an order of magnitude more sensitive to changes of fundamental frequency than to changes of other speech or music signal parameters.

For an arbitrary speech signal uttered by an unknown speaker, the fundamental frequency can vary over a range of almost four octaves (for male 50 to 800 Hz and for female 200 to 1400 Hz)

3.5.1 Basic Definitions of Pitch

There are three points()[?] of view for looking at a speech processing problem: the *production*, the *signal processing*, and the *perception* points of view. In the actual case of pitch determination the production point of view is oriented toward the generation of the excitation signal in the larynx; thus we will start from a time domain representation of the waveform as a train of laryngeal pulses.

If an algorithm is based on speech-production, it measures individual laryngeal excitation cycles or, if some averaging is performed, it determines the rate of vocal fold vibration. The signal processing point of view can be characterized in such a way that (quasi-)periodicity is observed in the signal and the task is just to extract the features that best represent this periodicity. The pertinent terms are *fundamental frequency* and *fundamental period*. The perception point of view leads to a frequency domain representation. In the technical literature the term pitch has consistently been used as a general name for all the terms mentioned before.

Defining the different representations of pitch, its reasonable to proceed from production to perception. So the basic definition based on speech production is as follows:

> *T_0 is defined as the elapsed time between two successive laryngeal pulses. Measurement starts at a well-specified point within the glottal cycle, preferably at the point of glottal closure or -if the glottis does not close completely- at the point where the glottal area reaches its minimum.* (1)

Pitch determination algorithms(PDAs) that obey this definition will be able to locate the point of glottal closure to delimit individual laryngeal excitation cycles. This task goes far beyond the scope of ordinary pitch determination.

> *T_0 is defined as the elapsed time between two successive laryngeal pulses. Measurement starts at an arbitrary point within the glottal cycle. Which*

point that is depends on the individual method, but for a PDA this point is always located at the same position within the glottal cycle. *(2)*

Ordinary time domain PDAs follow this definition. The reference point is not necessarily the point of glottal closure.

T_0 is defined as the average length of several periods , i.e., as the average elapsed time between a small number of successive laryngeal cycles. How the averaging is performed and how many periods are involved are matters of the individual method. *(3a)*

This is the standard definition of T_0 for any PDA that applies stationary term analysis, including the implementations of frequency domain . Well-known autocorrelation method follow this definition. The corresponding frequency domain definition is as follows.

F_0 is defined as the fundamental frequency of an (approximate) harmonic pattern in the (short-term) spectral representation of the signal. It depends on the particular method whether F_0 is calculated as the frequency of a certain harmonic divided by the respective harmonic number, as the frequency difference between adjacent spectral peaks, or as the greatest common divisor of the frequencies of the individual harmonics. *(3b)*

The perception point of view of the problem leads to a different definition of pitch. Pitch perception happens in the frequency domain. According to the existing theories.

F_0 is defined as the frequency of the sinusoid that evokes the same perceived pitch. *(4)*

PDAs that claim to be perception oriented enter the frequency domain in a manner similar to that in frequency domain definition (3b), i.e., by a standard short-term transformation such as Discrete Fourier Transform(DFT) with previous windowing of the signal.

3.5.2 Pitch Determination Methods

The existing PDA principles can be split up into two gross categories

1. Time Domain Methods: These methods will measure T_0 according to one of definitions (1) through (2).

2. Frequency Domain Methods: In all other cases, somewhere time domain is left; in this case F_0 or T_0 is determined according to definition (3a),(3b), or (4).

Time-domain methods using auto-correlation functions or difference norms are most popular and robust. In this thesis auto-correlation method is used and it is verified that it has negligible error for slow-varying, clean, monophonic singing voice signals.

3.5.3 The Autocorrelation Method

The voice signal is split up into a series of frames; an individual frame is obtained by taking a limited number of consecutive samples of the signal $x(n)$ from the starting point, $n = q - K + 1$, to the ending point, $n = q$ i.e. using rectangular window. The frame length, K, is chosen short enough so that the parameters to be measured can be assumed approximately constant within the frame. On the other hand, K must be large enough to guarantee that the parameter remains measurable. Frame thus requires two or three complete periods at least.

The autocorrelation function of a discrete time signal is defined as

$$\Phi(k) = \sum_{m=-\infty}^{\infty} x(m)x(m+k) \tag{3.1}$$

If the signal is periodic with period P samples, then the autocorrelation function is also periodic with the same period i.e.

$$\Phi(k) = \Phi(k+P) \tag{3.2}$$

Other important properties of the autocorrelation function are:

1. It is an even function; i.e., $\Phi(k) = \Phi(-k)$.

2. It attains its maximum value at $k = o$; i.e.,$|\Phi(k)| \leq \Phi(0)$ for all k.

3. The quantity $\Phi(0)$ is equal to the energy for deterministic signals or the average power for random, periodic signals.

If we consider equation (3.2) together with properties (1) and (2), we see that for periodic signals, the autocorrelation function attains a maximum at samples $0, \pm P, \pm 2P \ldots$. That is, regardless of the time origin of the signal, the period can be estimated by finding the location o the first maximum in the autocorrelation function. This property makes the autocorrelation function an attractive basis for estimating periodicities in all sorts of signals, including speech.

Let us define the short-time autocorrelation as

$$R_n(k) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)x(m+k)w(n-k-m) \tag{3.3}$$

This equation can be interpreted as follows: first a segment of speech is selected by multiplication by the window; then the autocorrelation definition (3.1) is applied to the windowed segment of speech. we can rewrite (3.3) in the form

$$R_n(k) = \sum_{m=-\infty}^{\infty} x(n+m)w'(m)x(n+m+k)w'(k+m) \tag{3.4}$$

where $w'(n) = w(-n)$. Above equation states that the time origin of the input sequence is effectively shifted to sample n, whereupon it is multiplied by a window w' to select a short segment of speech. If the window w' is of finite duration then the resulting sequence, $x(n+m)w'(n)$ will be of finite duration and (3.4) becomes

$$R_n(k) = \sum_{m=0}^{N-1-k} [x(n+m)w'(m)][x(n+m+k)w'(k+m)] \tag{3.5}$$

The calculation of the autocorrelation for the entire range of lag can be done using (3.5). Using these values of autocorrelation, it is possible to find the value of lag associated with the highest autocorrelation representing the pitch period estimate, since, in theory autocorrelation is maximized when the lag is equal to the pitch period.

3.6 Harmonic Analysis

Most of the research in the frequency domain analysis has been based around the Fourier transform. Specifically, the Discrete Fourier Transform (DFT) and its efficient Fast Fourier Transform (FFT) version comprise the backbone of many studies in the frequency domain.

The waveform of voiced speech (e.g.vowels) is often nearly periodic. The periodicity in turn is shown in the Fourier-transform s.t. its DFT is harmonic, which is most of its energy is in the fundamental frequency f_0 and its multiples $2f_0, 3f_0, 4f_0$,and so on. The DFT is defined as:

$$X[k] = \sum_{m=0}^{N-1} x[n]e^{-j2\pi kn/N} \tag{3.6}$$

Where,

X[k] is a complex number with magnitude and phase components at frequency bin k,

n is discrete time index,

x[n] is sampled input signal,

N is length of DFT(usually equal to the length of the window).

When applying the DFT, care should be taken in the selection of parameters as a compromise between frequency and time resolution always exists: increased time reso-

lution (transitory characteristic) leads to a degraded frequency resolution resulting in frequency smearing. One way to elude this problem is using the Short-Time Fourier Transform (STFT) which is defined as

$$X[k] = \sum_{m=0}^{N-1} w[k - mD]x[n]e^{-j2\pi kn/N} \tag{3.7}$$

The STFT shown in equation (4.1) can be simply thought of as windowing a signal, but rather than advancing or hopping the starting point of the signal x[n] by the window size, windows are overlapped and advanced depending on the overlap length. In effect, this lessens some of the degradation of time-frequency smearing and is applied in most DFT-based spectral analysis practices. Also different window types such as rectangular, Hamming, Blackwell, and others exist for extracting a slice of a signal. Each window has its own particular shape which determines the side-lobe characteristics.

Location and strengths of harmonics can be obtained using the peaks in the power spectrum obtained using STFT. Detailed procedure is as follows:

Step 1: The input signal is decomposed into overlapped frames with hop size half the frame size. The hop size is the time interval between the centers of two adjacent frames. Each frame is windowed by a 100 ms long hanning window.

Step 2: Perform the short-time Fourier transform on each frame.

Step 3: In this step, we want to extract the partials of each frame which is done using harmonic analysis algorithm.

3.6.1 Harmonic Analysis Algorithm

The strength and locations of the harmonics is important aspect of the spectral analysis. Harmonic analysis pertains primarily to feature extraction for pitched speech or music signals. As we shall see in this section, armed with the information regarding harmonics and their behavior, features such as spectral centroid, tristimulus characteristics, etc, can be extracted.

Following Harmonic analysis algorithm is developed to find out harmonics, inharmonics, and their power levels in the power spectrum obtained using the STFT.

1. Most salient components(candidate harmonics) of the power spectrum are deter-

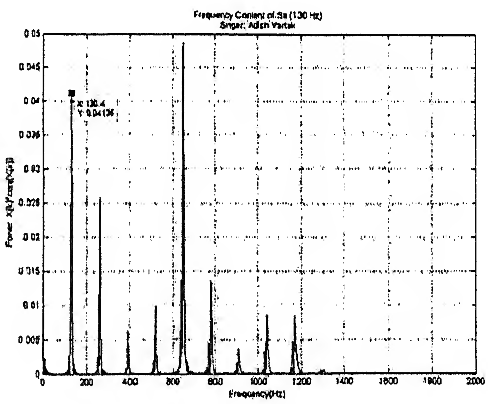


Figure 3.1: Power Spectrum of Note Sa(Steady State), Singer: Adish Vartak

mined using the strongest component present in the spectrum.

Let P_{max}=Power in the strongest component(harmonic) of the spectrum

If any,

$$Component \geq XP_{max} \tag{3.8}$$

where, X=Detection Threshold for Components, e.g. In this thesis $X = 0.02$ (which is 2% of P_{max}) is used. We can set the value of this parameter depending on the noise present in the spectrum and the accuracy required in the detection of the harmonics.

2. The harmonic lengths (distance between components(harmonics))are determined using following equation:

$$(HarmonicLength) = abs(CurrentComponent - LastComponent) \tag{3.9}$$

thus harmonic length vector is prepared.

3. Fundamental Frequency Determination:

i) All these harmonic lengths are compared and the length which is occurred maximum number of times is compared with the first component of the spectrum and

if,

$$abs(Harmoniclength - Firstcomponent) \leq Deviation, \tag{3.10}$$

then,

$$FirstComponent = FirstHarmonic$$

Where deviation is the deviation from the ideal values of the harmonics. We can set the value of this parameter as per required accuracy.

else if,

$$abs(Harmoniclength - Secondcomponent) \leq Deviation, \tag{3.11}$$

then $SecondComponent = FirstHarmonic$ and first component is inharmonic.

Above step is repeated unless we get first harmonic i.e. fundamental f_0.

ii) If the number of components in the power spectrum are less i.e. 1,or 2 following procedure is used:

If only one component is present, that component is taken as fundamental, when two components are present then these are checked for multiplicity, if multiple, are taken as fundamental and second harmonic otherwise first is taken as first harmonic and second component is taken as inharmonic.

4. Once the fundamental frequency is determined, remaining components are checked for respective harmonics using following equation.

if,

$$abs(SecondComponent - FirstHarmonic) \leq Deviation, \tag{3.12}$$

then

$$SecondComponent = SecondHarmonic,$$

else,

second component is inharmonic.

5. Similarly all components are checked for respective harmonics and harmonic vector is prepared.

6. Using power spectrum, Power level of the respective harmonics and inharmonics is determined.

3.7 Representation of the Timbre

It is important to note that timbre is a perceptual quality of a sound, much like colour is a perceptual quality of light. Hence the dimensions of the timbre are the parameters which our ears translate to information about the quality of a particular sound. Pollard

and Jansson([13]) suggested that the relative weighings of three bands of partials in an acoustic signals may convey sufficient information for the brain to make an informed evaluation of timbre. The inspiration for this proposal was largely due to considering the process of colour detection in humans, where only three types of receptors detect incoming light.

3.7.1 The Tristimulus Theory

The tristimulus theory breaks down the harmonic partials into three separate bands:

1. The fundamental frequency, f_o.

2. Mid-frequency partials(identified as 2nd to 4th).

3. High-frequency partials (5^{th} to n^{th} harmonic, n being the highest significant partial).

The total loudness (N) is normalized to unity and is calculated as the sum of the loudness values of the three bands.i.e., The tristimulus is defined by the following three equations.

$$z = Tristimulus1 = \frac{H[1]}{\sum\limits_{k=1}^{N} H[k]} \tag{3.13}$$

$$y = Tristimulus2 = \frac{H[2] + H[3] + H[4]}{\sum\limits_{k=1}^{N} H[k]} \tag{3.14}$$

$$x = Tristimulus3 = \frac{\sum\limits_{k=5}^{N} H[k]}{\sum\limits_{k=1}^{N} H[k]} \tag{3.15}$$

Where H[N] is the upper most harmonic, and k=1 refers to the fundamental component.

The relative intensities of these three bands can be plotted on a two-dimensional triangular diagram where each corner represents a total concentration of a energy in a

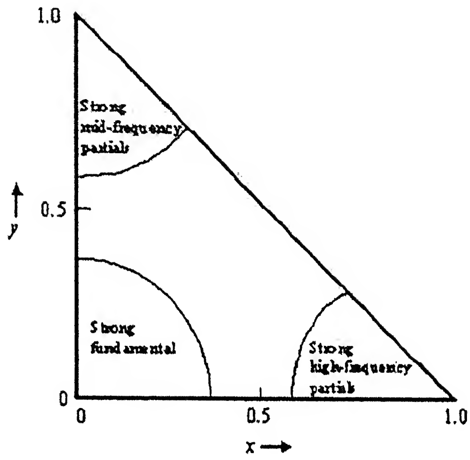


Figure 3.2: Basic Layout of Tristimulus Diagram

particular band. The layout of the tristimulus diagram is shown in figure. Plotting the mid and high frequency bands is enough to specify a point on the diagram because information about the loudness of the fundamental frequency can be inferred from the proximity of the point to the origin.

3.7.2 Reading Tristimulus Diagram

The lifetime of note is traced through the tristimulus diagram using the x and y dimensions. From these plots we can get information about musical quality of notes. As we

Sr.No.	Quality	Harmonic Structure
1	Hollowness	Lower odd harmonics dominates.
2	Nasality	Higher odd harmonics dominates.
3	Roughness	Higher harmonics or only odd or even harmonics dominate.
4	Dull,very soft	Single tone or, single tone plus Inharmonics.
5	Good musical quality	First six moderately loud.
6	Richness	Fundamental dominates.
7	Softness	Higher harmonics (¿6) should be absent.
8	Brightness	More number of harmonics, High Harmonic Spectral Centroid.

Table IV: Reading Tristimulus Diagram

have seen if first six harmonics are moderate then it represents good musical quality. So for a good quality note we expect comparable powers in mid and high band. Above table represents particular qualities and its structure in the power spectrum.

3.8 Spectral Centroid

The spectral centroid corresponds to a timbral feature that describes the brightness of a sound. This important feature has been elicited in the past ([?]). The spectral centroid can be thought of as the center of gravity for the frequency components of

a signal. It exists in many variations including its mean, standard deviation, square amplitudes, log amplitudes, and the harmonic centroid . The centroid, currently one of the MPEG-7 timbre descriptors, is defined as:

$$SpectralCentroid = \frac{\sum_{k=1}^{N} kf_0 P_H[k]}{\sum_{k=1}^{N} P_H[k]} \tag{3.16}$$

$P_H[k]$ is the Power in the K^{th} harmonic,

N is the total number of salient harmonics in the Power Spectrum.

Generally, it has been observed that sounds with dark qualities tend to have more low frequency content and those with brighter sound dominance in high frequency (Backus 1976) which can be inferred by the value of the centroid. It has also been suggested (Kendall, Carterette 1996) that the centroid be normalized in pitch hence making the spectral centroid a unit-less and relative measure since it is normalized by the fundamental frequency f_0. Some researchers have therefore included both the normalized and absolute versions of the centroid (Eronen, Klapuri 2000).

Chapter 4

Proposed Method and Results

4.1 Experimental Details

Singers were asked to select three tonics of their choice. Then they were asked to sing aaroh and avaroh with the selected tonic in akaar(aalap) i.e. without pronouncing any syllables, only using the sustained /aa/ vowel sound. The waveform of voiced speech (e.g. vowels) is often nearly periodic. The periodicity can be seen in the fourier-transform. i.e. its DFT is harmonic, which is most of its energy is on the fundamental frequency f_0 and its multiples $2f_0$, $3f_0$, $4f_0$ and so on.

These signals(slow varying, monophonic) were recorded using desktop computer and a microphone. The sampling frequency used was 16 KHz. Only the voice of singer was recorded. The notes were separated using Gold-Wave recording software. In this experiment several audio clips drawn from commercial recordings of professional singers (e.g. Lata Mangeshkar, Md. Rafi, etc.)are used for illustration of criterion, used for the tonic selection.

Every note is divided into three sections, rising phase, the steady-state(sustain) and

the decay . The steady-state corresponds to the portion of a note where the amplitude is stable and a constant pitch is observed. Using pitch plots transient and steady state parts of the notes were separated. It is also verified that the spectrum of the notes in steady state part is not changing very much. For all the notes these steady state portions were obtained and for further analysis power spectrums of these steady state portions were used.

4.2 Factors Influencing the Choice of Tonic

In the next few subsections we develop and illustrate criterion for the tonic selection. In the previous chapters we have discussed some fundamentals of Indian music and also we have reviewed some past and present research on timbre which is related to the quality of sound.

Ideally in Indian music notes are expected to be powerful and steady with very high value of intonation. These important factors are discussed and also from perception point of view some research on timbre is reviewed(e.g. Richness, Brightness, etc these type of qualities are discussed).

4.2.1 Intonation

For every note RMSE is calculated, which is defined as:

$$RMSE = \sqrt{E\{[\tilde{p}(n) - p(n)]^2\}} \tag{4.1}$$

Where,

$\tilde{p}(n)$ = Measured Pitch and,

$p(n)$ = Expected(Ideal) Pitch.

This represents deviation(in Hz) from the ideal or expected pitch of the note.

We asked singers about the name of the notes and the scale used in aaroh and avaroh. This information is used in RMSE calculations. Pt. Bhatkhande's scale is used in these calculations as told by the singers.

This is used as a measure for the intonation(deviation from the ideal or expected pitch). From this we can get information about whether the given note is reached or not and the error between the rendered pitch and the expected pitch.

Intonation is considered to be very important aspect in Indian music(1). Proper intervals or ratios of each note with the reference or basic note are fixed as given in table 1.3. Ideally pitch of the note should coincide with the pitch given in table 1.3.

4.2.2 Richness

Richness is determined by the power in the fundamental harmonic. When deciding the tonic of singer, fundamental harmonic plays important role because Indian music is based on the tonic or Sa. As we know, a raga can be identified only if the Sa note pitch(tonic) is identified as the other notes in the scale are related to the basic tonic. So the fundamental harmonic of this note should not be too weak rather it should be powerful because the pitch perceived is strongly correlated with the fundamental harmonic(p1) and also when the fundamental predominates, the quality of the tone is found to be rich, but when the fundamental was weak, the quality was poor(3).

In analysis of Indian music singer's voices i.e. Lata Mangeshkar, Md. Rafi's voice, it is found that the fundamental harmonic in almost all the notes is dominating. This supports our criterion that this note should be rich i.e. fundamental harmonic should

not be too weak.

4.2.3 Brightness

Brightness is related to the number of harmonics present in the note. More harmonious notes are preferred. Harmonic Spectral centroid, which is defined as

$$SpectralCentroid = \frac{\sum_{k=1}^{N} k f_0 P_H[k]}{\sum_{k=1}^{N} P_H[k]}$$

is calculated for every note. This is used as a measure for brightness of the note.

In analysis of Indian music singer's voices i.e. Lata Mangeshkar, Md. Rafi's voice, it is found that these voices are reasonably bright and also Helmholtz found musical notes which were accompanied by a moderately loud series of the lower partial tones up to about the sixth were more harmonious and musical, while they were at the same time sweet and soft if the higher partials were absent(Helm.).

4.2.4 Power

Indian music singing requires a voice which sounds pleasant and to be heard with power in all the three registers or octaves so note should not be too weak [20]. Specifically in lower octave it is observed that the power is becoming considerably low. Using this criterion along with other criterions, we can get lower limit of the singer's voice.

4.2.5 Steadiness

It is observed that the pitch of the notes is varying around steady state value. In Indian music, steady notes are preferred, so notes should not deviate too much from

the steady state value. This can be observed from the pitch plots.

4.2.6 Relative Strengths of Odd/Even Harmonics

If only odd or even harmonics are dominating in the note, then the pitch perceived will be an octave higher than the rendered pitch. Ideally, according to Jeans, odd harmonics should not dominate, and at the same time they should not be too weak. If weak, perceived pitch will be quite high because it will be decided only by even harmonics [1, 2].

4.3 Analysis of Voices of Singers

In the previous section we have seen the factors influencing the choice of tonic. Some of the factors are illustrated using classical theory of timbre(quality) along with some present research in timbre and other factors are the criterion laid down by Indian musicologists for rendering the musical notes. It is important to examine these factors for good Indian music singers.

Aalap portions of various hindi songs and ragas of following singers are analyzed to examine various features:

1. Lata Mangeshkar

2. Md. Rafi

In the following subsection we will analyze Lata Mangeshkar's voice and in subsequent subsection we will analyze Md. Rafi's voice.

4.3.1 Lata Mangeshkar's Voice

Aalap portion of bhajan 'Jai Ram....' is analyzed. Notes were separated using Gold-Wave software. For all the notes pitch plots were obtained and from pitch plots constant pitch part of the note was selected. Using constant pitch part of the notes power spectrums of these notes were obtained.

4.3.1.1 Pitch Plots

Pitch plots of steady state notes are shown in fig. It can be observed in this plot that

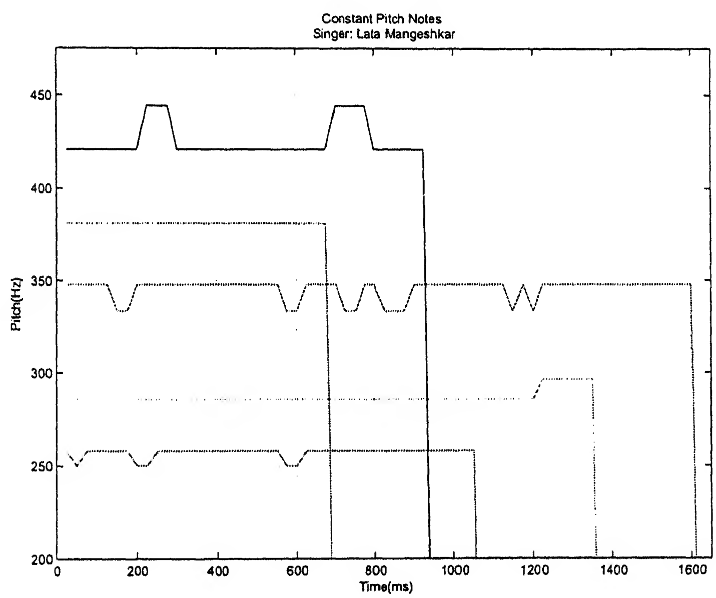


Figure 4.1: Pitch Plot: Lata Mangeshkar's Voice

the pitch of some of the notes is not constant through out the note. Some variations can be observed. For analysis, only steady state portions of the notes were selected, where pitch of the note was almost constant.

4.3.1.2 Power Spectrum

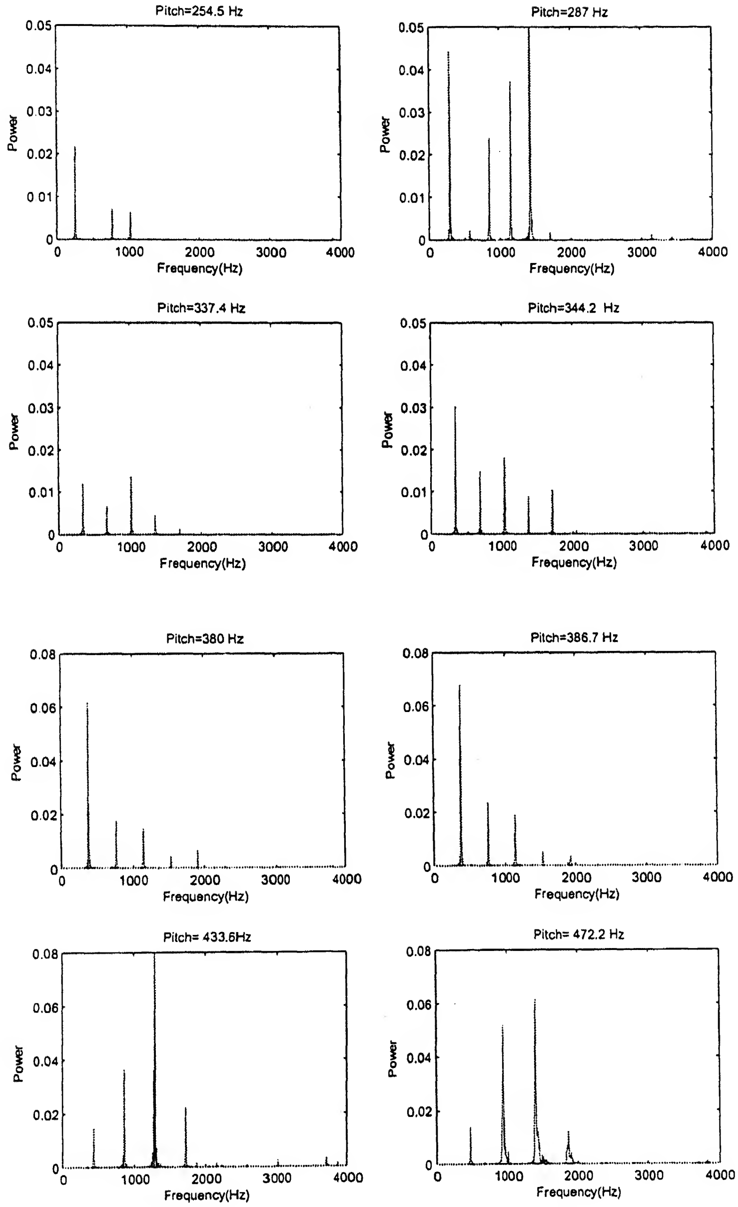


Figure 4.2: Power Spectrum: Lata Mangeshkar's Voice

Power spectrums of eight constant pitch notes are obtained which are shown below. Length of all the notes was taken same so that we can compare power levels of the notes. As can be seen harmonics are clearly visible and inharmonics are not observed in these power spectrums.

Notes are selected such that they span complete octave and we get characteristics of voice over complete octave. Power spectrums of these notes are shown in figure. Two main things that can be observed in these spectrums are:

1. Strength of fundamental harmonic.
2. Number of harmonics present and their relative strengths.

Regarding the first factor, it can be observed that in almost all the notes fundamental harmonic is dominating. The only exceptions are the two highest notes in the octave where second and third harmonics are dominating but it can be noted that even at this high pitch (octave higher than the lowest note) power in the fundamental harmonic is considerable.

Regarding the second factor, in almost all the notes most of the power is carried by first five harmonics. Even at the high pitch notes(octave higher) four strong harmonics are present and power is not carried by single harmonic but is distributed among the four harmonics with the dominance of second and third harmonics.

So according to classical timbre theory, we can conclude that this voice is rich and bright which is evident from above two points i.e. fundamental harmonic and number of harmonics present in the notes.

Also the second important thing we note is that, tonic or Sa of this singer will be the one of the first six notes because normally, for female singers tonic range is 200 Hz

to 260 Hz, in exceptional cases tonic can go outside this range [7]. In the first six notes

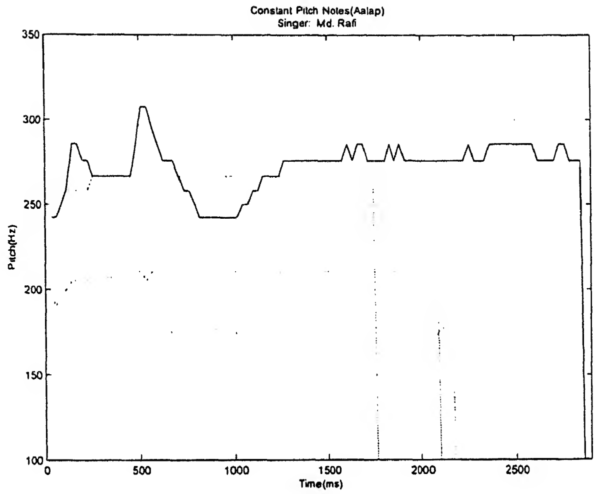


Figure 4.3: Pitch Plot: Md. Rafi's Voice

we can note the harmonic structure, fundamental is strongest with other harmonic amplitudes decreasing with number.

4.3.2 Md. Rafi's Voice

Aalap portion of 'Duniya Na Bhaye Mohe' song(Basant Bahar) which is in raga todi is selected to analyze Md. Rafi's voice. Pitch plots of constant pitch notes in the aalap are shown. We can see in these plots that, during the rising phase pitch is varying from some initial value and it is taking some time to attain the steady state pitch. Steady state portions of these notes i.e. portion where pitches of these notes are constant are selected for the analysis.

4.3.2.1 Power Spectrum

Power spectrum of steady state portion of the note is shown in the diagram. It can be observed that the number of harmonics at low pitches are around ten. Fundamental

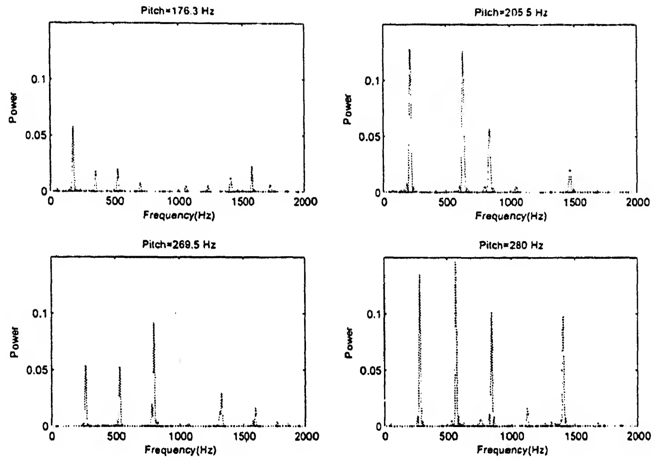


Figure 4.4: Power Spectrum: Md. Rafi's Voice

harmonic is dominating in almost all the notes. Second and third harmonics are also powerful.

4.4 Selection of Tonic

Singers were asked to select three tonics of their choices(where they think they are comfortable). Then they were asked to sing aaroh and avaroh with the selected tonic in akaar(aalap). From these three tonics we can get his voice range and by analyzing these notes for the various features such as RMSE, Harmonic Centroid, Power in the fundamental harmonic, total power, and using tristimulus diagrams, we can fix their

tonic. In the following sections we will analyze voices of the two singers Adish Vartak and Rajendra Singh.

The following notations are used to describe notes.

Lower notes are written in lower case and upper notes in upper case. Thus,

1. Shuddh notes are notated as S, R, G, m, P, D, N
2. Komal notes are notated as r, g, d, n
3. Teevra Ma is notated as M

All notes belong to madhya-saptak by default. Notes of mandra-saptak are preceded by ‘ sign, and notes of taar-saptak are succeeded by ’ sign. For instance ‘N means Ni of mandra-saptak, and S’ means Sa of taar-saptak.

4.5 Selection of tonic: Adish Vartak

Adish selected three tonics which were 120, 130, 140 Hz. His notes in the aaroh were: Sa, re, Ga, ma, Pa, dha, Ni, Sa’, re’, Ga’, ma’
where, Single closing quote(’) after the note name indicates the note of the higher octave, and note name in the small letters indicate komal(flat) note, Note name in the capital letter indicates shuddha note. Teevra(Sharp) Ma is notated as M.

In the avaroh his notes were: Sa’, Ni, dha, Pa, ma, Ga, re, Sa, ‘Ni, ‘dha, ‘Pa, ‘ma where, Single opening quote (‘) before the note name indicates the note of the lower octave. Once tonic(Sa) is selected, using Bhatkhande’s scale or intervals, pitches of the remaining notes are calculated. We will analyze all these notes first to find out his voice range and then we will select tonic according to his voice range and our criterion for tonic selection.

Steady state parts of the notes were selected using pitch plots. Power spectrums of all these notes are shown in the figure. These power spectrums are analyzed for the various factors discussed in the previous sections.

4.5.1 Singer's Voice Range

In this subsection we will analyze notes in the aaroh and avaroh to find out singer's voice range. Analyzing aaroh we can get higher note and analyzing avaroh we can get lower limit of voice range. Various parameters are calculated for all the notes as shown in the tables.

4.5.1.1 Analysis of Aaroh

All the notes in the aaroh with three Sa are analyzed. Pitch plots, power spectrums are obtained as shown in the figures. From these power spectrums following parameters are obtained.

Pitch of the Note: Pitch of the note is the frequency of the fundamental harmonic.

As can be seen in the table 4.6.3 and power spectrums, when the tonic is 120 Hz, highest note is 328.61 Hz with the parameter values RMSE 2.51, Harmonic Centroid 2.35, Fundamental Harmonic Power 11.34% of total power of the note, which indicates that this note is reached comfortably with good harmonic structure.

RMSE: Root Mean Square Error is the measure for intonation. This indicates the error between ideal or expected pitch and the actual pitch reached. As can be seen from the tables, maximum error in the intonation is at the highest note ma' with the tonic 141 Hz, which is 12.56 Hz.

Power in the Fundamental Harmonic: As we have seen richness is determined by

Sr. No.	Note Name	Expected Pitch(Hz)	Observed Pitch(f_1)	RMSE	Harmonic Centroid($*f_1$)	Power in f_1 (%)	Avg. Power in Note
1	S	121.58	121.58	1.2998	3.0596	36.638	1.7742
2	r	128.09	130.86	2.4313	3.1698	31.434	1.6114
3	G	151.98	155.27	2.9182	2.9866	42.568	1.5035
4	m	162.11	164.06	1.6819	3.1315	31.95	2.8007
5	P	182.37	181.64	1.9703	3.296	22.505	4.6683
6	d	196.1	191.41	4.5042	2.9432	28.295	4.7148
7	N	227.97	231.45	3.7258	2.8647	12.568	17.605
8	S'	243.16	245.12	2.5373	2.5459	29.743	12.914
9	r'	256.17	260.74	3.6817	2.5516	24.343	16.162
10	G'	303.96	308.11	3.9088	2.6913	16.962	19.08
11	**m'**	**324.22**	**328.61**	**2.5186**	**2.3543**	**11.347**	**22.474**

Table 4.1: **Aaroh 120 adish**

the power in the fundamental harmonic. Weakest fundamental harmonic is observed at the highest note i.e. at the note ma' in aaroh with tonic 141. At this note 2.9 % of the total power is carried by the fundamental harmonic.

Harmonic Centroid: Harmonic centroid represents center of gravity of the spectrum. This is used as a measure for the brightness of the note. This indicates the number of

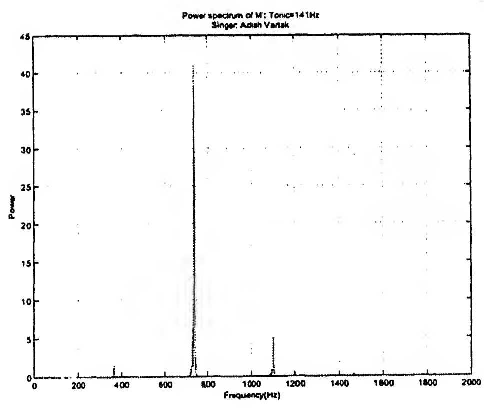


Figure 4.5: Power Spectrum of m': Sa= 141 Hz

Sr. No.	Note Name	Expected Pitch(Hz)	Observed Pitch(f_1)	RMSE	Harmonic Centroid($*f_1$)	Power in f_1 (%)	Avg. Power in Note
1	S	131.35	131.35	0.89652	2.8027	40.446	0.47943
2	r	138.37	137.7	1.3272	3.7868	23.758	0.61277
3	G	164.18	164.06	1.0524	3.4222	29.468	0.73695
4	m	175.13	172.85	3.0402	3.3515	23.662	1.382
5	P	197.02	195.31	2.8783	3.1149	20.833	1.6622
6	d	211.85	205.57	7.6327	3.0654	23.29	2.7027
7	N	246.28	243.16	3.7545	3.0387	7.0839	11.242
8	S'	262.7	259.28	4.3644	2.5632	31.185	3.0134
9	r'	276.75	273.44	4.8654	2.5985	21.725	5.135
10	G'	328.37	325.68	4.3702	2.6639	6.9737	13.143
11	**m'**	**350.26**	**344.73**	**9.1747**	**2.3316**	**6.5315**	**13.217**

Table 4.2: **Aaroh 131 adish**

Sr. No.	Note Name	Expected Pitch(Hz)	Observed Pitch(f_1)	RMSE	Harmonic Centroid($*f_1$)	Power in f_1 (%)	Avg. Power in Note
1	S	141.11	141.11	1.3552	3.8811	22.013	1.7602
2	r	148.66	147.46	1.5303	4.4762	13.326	2.3686
3	G	176.39	174.32	2.2678	3.802	13.628	3.5346
4	m	188.15	185.06	4	3.5202	16.231	4.0491
5	P	211.67	208.98	3.9318	3.4144	16.564	6.6119
6	d	227.6	216.8	10.157	3.0337	24.809	4.5189
7	N	264.59	262.7	2.45	2.7514	14.801	10.418
8	S'	282.23	276.37	6.4165	2.5963	22.577	7.2965
9	r'	297.33	291.02	7.7309	2.4603	22.538	8.9628
10	G'	352.78	346.68	6.7476	2.2404	9.8804	14.145
11	**m'**	**376.3**	**367.19**	**12.565**	**2.0948**	**2.944**	**48.443**

Table 4.3: **Aaroh 141 adish**

harmonics present in the spectrum and the center of gravity of the spectrum .

Total Power in the Note: It is observed that in the middle and higher octaves, notes are strong and in the lower octave the notes are weak. In the selection of the lowest note this measure is used i.e. in the analysis of avaroh this measure can be used to find out the lowest note. In aaroh all the notes are reasonably powerful.

We can observe in the table 4.3 that the highest note in this table is m'. This note

is ideally expected at 376.3 Hz but the singer is reaching at the 367.19 Hz, RMS error is 12.565 which is considerably large.

If we observe the harmonic structure of this note(fig.4.5), we can observe that the fundamental harmonic is very weak as compared to the second harmonic and also harmonic centroid is minimum at this note. This note according to timbre theory is not rich and bright and does not represent perceptually good note . This note is at the pitch of 367 Hz, so due to above factors we consider his highest note to be less than this pitch.

If we observe the lower note G' in the same table i.e. at the pitch of 346.68 Hz, for this note we can see the calculated parameters are well within acceptable limits, which can also be observed from the power spectrum. So we can conclude that his highest note is around 346 Hz.

Sr. No.	Note Name	Expected Pitch(Hz)	Observed Pitch(f_1)	RMSE	Harmonic Centroid($*f_1$)	Power in f_1 (%)	Avg. Power in Note
2	N	244.45	242.68	2.2462	3.0927	11.642	3.5528
3	d	210.28	202.64	7.7221	3.4179	11.949	5.0823
4	P	195.56	192.38	3.4443	3.0961	18.871	1.9067
5	m	173.83	170.9	3.6544	3.5493	17.742	2.2398
6	G	162.96	161.13	1.9214	3.7417	12.984	2.6533
7	r	137.35	136.72	1.9884	4.3222	19.066	0.68691
8	S	130.37	130.37	1.2278	2.4059	50.744	0.72236
9	'N	122.22	123.05	1.4204	3.959	20.838	1.037
10	'd	105.14	104	2.0382	4.2795	17.865	0.76274
11	'P	97.778	96.68	0.95149	3.9959	21.084	0.43621
12	'm	**86.914**	**85.938**	**1.0902**	**3.2589**	**33.971**	**0.10659**

Table 4.4: **Avaroh 131 adish**

4.5.1.2 Analysis of Avaroh

All the notes in the avaroh with three Sa are analyzed. Pitch plots, power spectrums are obtained as shown in the figures. From these power spectrums, tables, our observations are:

Pitch of the Note: As can be seen from the table, lowest note reached is 83.49 Hz, with the tonic 124 Hz.

Sr. No.	Note Name	Expected Pitch(Hz)	Observed Pitch(f_1)	RMSE	Harmonic Centroid($*f_1$)	Power in f_1 (%)	Avg. Power in Note
1	S'	249.02	248.05	1.9809	2.9752	16.16	10.704
2	N	233.46	234.38	1.8481	2.7539	20.459	10.079
3	d	200.83	198.24	2.6928	2.9235	23.915	4.1834
4	P	186.77	185.55	1.9137	2.6199	32.985	3.2803
5	m	166.02	166.02	1.0579	3.3033	23.049	2.8596
6	G	155.64	154.79	1.2734	3.7082	20.561	3.0362
7	r	131.17	129.88	2.3194	3.0604	37.72	1.1555
8	S	124.51	124.51	1.7704	2.6363	47.678	0.50045
9	'N	116.73	116.21	1.2291	3.95	26.065	0.89864
10	'd	100.41	98.633	1.8048	3.8323	25.695	0.49719
11	'P	93.384	94.238	1.453	4.0926	20.124	0.40525
12	**'m**	**83.008**	**83.496**	**1.2042**	**3.6927**	**27.364**	**0.19308**

Table 4.5: **Avaroh 124 adish**

Power in the Fundamental Harmonic: As we have seen richness is determined by the power in the fundamental harmonic. At this note 27.36 % of the total power is carried by the fundamental harmonic.

Harmonic Centroid: Harmonic centroid of this note is 3.69. Many harmonics can be observed in the power spectrum, which indicates good musical quality of the note.

Total Power in the Note: It is observed that in the middle and higher octaves,

Sr. No.	Note Name	Expected Pitch(Hz)	Observed Pitch(f_1)	RMSE	Harmonic Centroid($*f_1$)	Power in f_1 (%)	Avg. Power in Note
1	S'	271.48	273.93	4.0312	2.9032	14.225	8.3067
2	N	254.52	257.81	3.27	2.5884	18.775	8.1387
3	d	218.94	219.24	1.6531	3.0475	11.874	10.945
4	P	203.61	205.57	3.2076	3.389	11.385	8.5632
5	m	180.99	183.59	3.0263	3.0734	22.762	3.3544
6	G	169.68	175.29	4.1229	3.4553	20.769	3.2239
7	r	143	144.53	2.0561	4.068	17.002	2.2214
8	S	135.74	135.74	1.1382	4.3458	13.47	1.831
9	'N	127.26	127.44	0.76145	4.8899	12.606	2.1405
10	'd	109.47	109.86	0.82733	4.5558	13.111	1.1016
11	'P	101.81	103.52	1.5984	5.2129	10.216	1.1705
12	**'m**	**90.495**	**92.773**	**1.2665**	**4.3723**	**19.766**	**0.30157**

Table 4.6: **Avaroh 141 adish**

notes are strong and in the lower octave the notes are weak. In the selection of the lowest note this measure is important. It can be noted that the power level of this note is reasonable according to the singer.

In this note good harmonic structure is present. According to the comfortableness of the singer, we can conclude that this is his lowest note(84 Hz).

4.5.1.3 Singer's Voice Range

We can note from the previous two subsections that the voice range of this singer is:

Lowest Note is around 84 Hz and,

Highest Note is around 346 Hz.

4.5.2 Tristimulus Analysis

Effect of tonic on spectrums of various notes can be studied using tristimulus diagrams. As we know in the tristimulus diagrams whole spectrum is divided into three bands.

Only two bands on 2-D diagrams are plotted. Information about third band can be obtained easily due to the normalization of these three bands.

4.5.2.1 Comparison of Aaroh with Three Tonics

We can compare spectral range of aaroh with various tonics. Using classical theory of timbre we can find out the tonic, which has good spectral contents. We will compare aaroh with three tonics, as shown in the following figure. It can be observed in the

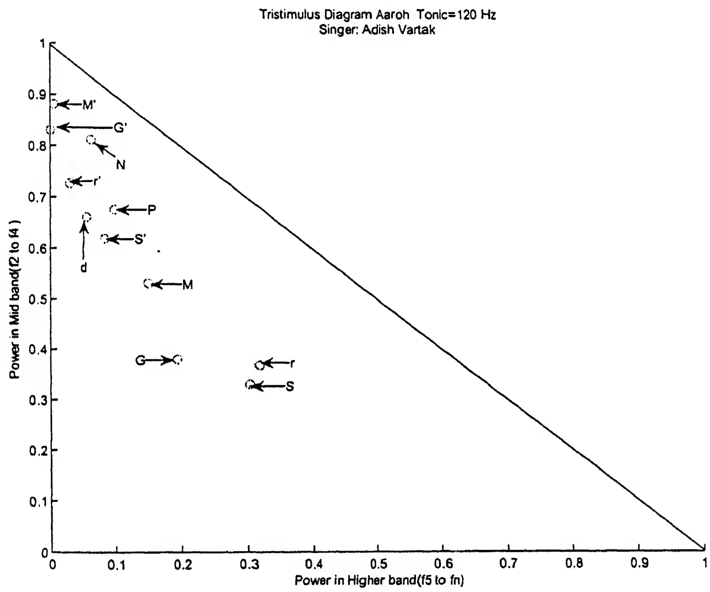


Figure 4.6: Spectral Variation: Aroh tonic= 120Hz, Singer: Adish Vartak

figure with tonic 120 Hz, that the first four notes S, r, G and M are at the central part of the diagram which indicates, In these notes strengths of fundamental, mid band and higher band are comparable which according to classical timbre theory indicates good musical quality. At the higher notes(notes in the higher octave) i.e. r', G', M' we

can observe that these are clustering towards upper corner indicating decrease in the strengths of fundamental and higher band, i.e. we can note that as pitch is increased, notes are leaving the central part of the diagram and clustering towards upper corner which indicates the deterioration of the quality of these notes. We can observe in this

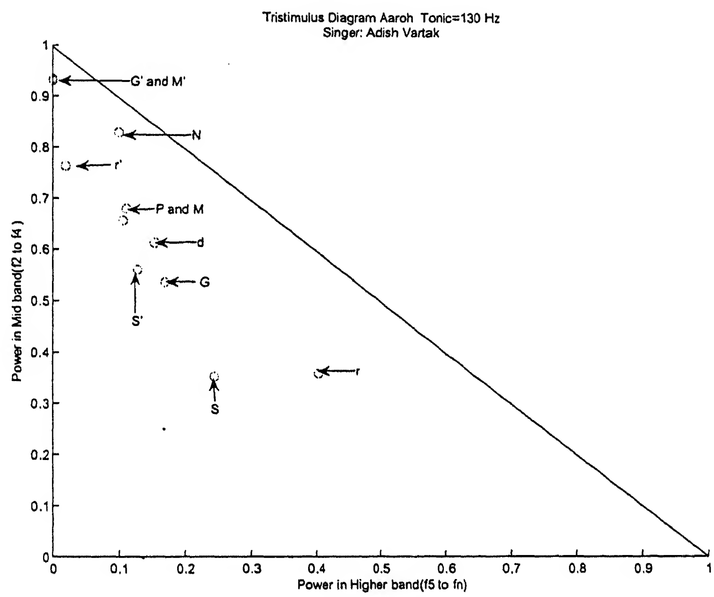


Figure 4.7: Spectral Variation: Aroh tonic= 130Hz, Singer: Adish Vartak

tristimulus diagram (tonic 130Hz) that, in the last two notes of the higher octave i.e. G' and m' higher harmonics are absent and also fundamental is considerably weak which indicates further deterioration in the musical quality of these notes.

We can observe in the tristimulus diagram (tonic 140 Hz) that cluster of all the notes is shifted towards upper corner except notes S and r. In the notes S and r clear dominance of higher harmonics is visible in the diagram and all the higher octave notes i.e. S', r', G', M' are on the y-axis which indicates higher harmonics are absent.

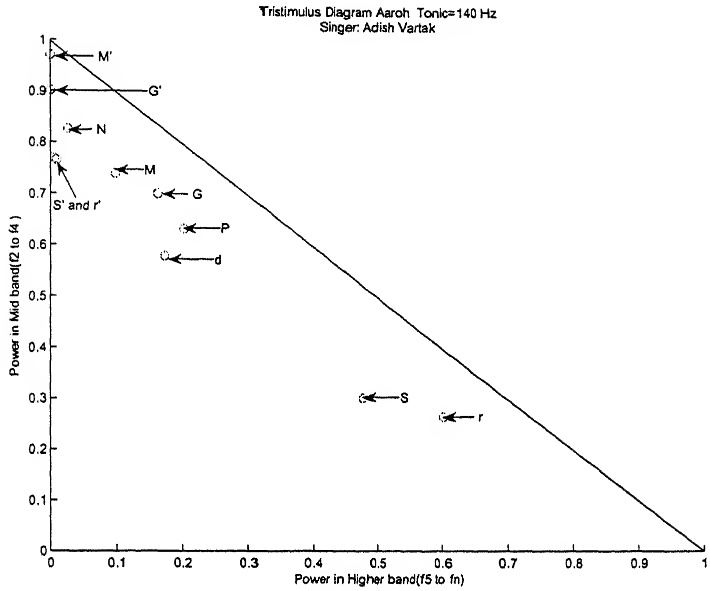


Figure 4.8: Spectral Variation: Aroh tonic= 140Hz, Singer: Adish Vartak

Clustering in the upper corner indicates weak fundamental and higher harmonics. We can observe in the note m', fundamental and higher harmonics are almost absent. This demands critical examination of this note.

4.5.2.2 Comparison of Avaroh with Three Tonics

As we can see in the tristimulus diagram, all the notes of the avaroh with 124 Hz are in the central part of the diagram which indicates good musical quality.

We can note that here singer has used 124.5 Hz tonic instead of 120 Hz. Now, in the avaroh with 130.4 Hz

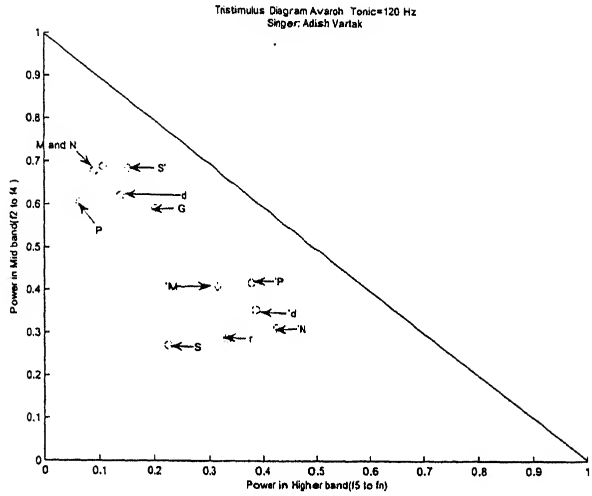


Figure 4.9: Spectral Variation: Avaroh tonic= 124.5Hz, Singer: Adish Vartak

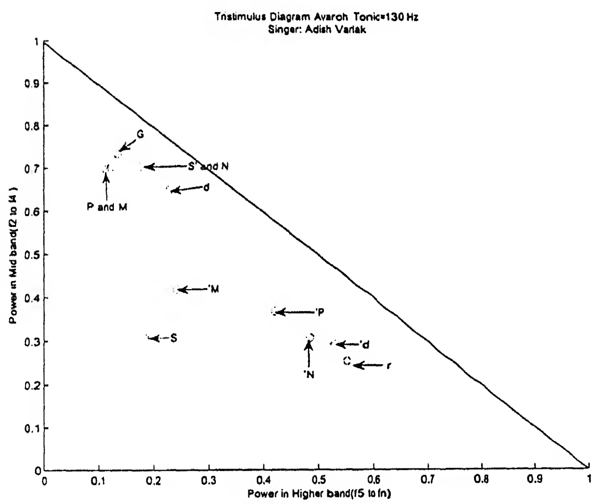


Figure 4.10: Spectral Variation: Avaroh tonic= 130Hz, Singer: Adish Vartak

We can see in fig. 4.11, a slightly spread structure. Notes are leaving the central part. Middle octave notes are moving towards upper corner and lower octave notes

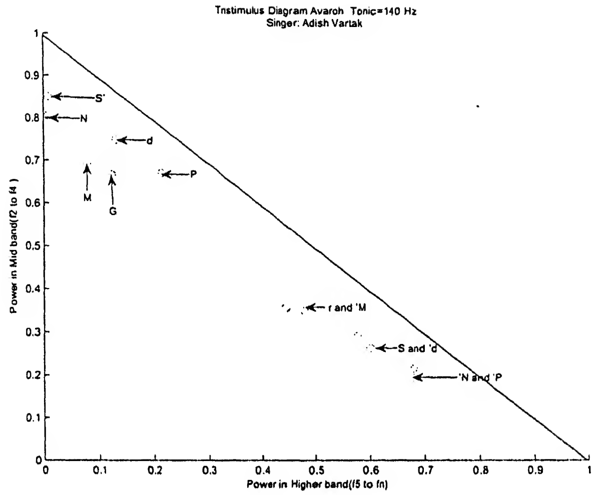


Figure 4.11: Spectral Variation: Avaroh tonic= 140Hz, Singer: Adish Vartak

are clustering towards right corner indicating dominance of higher harmonics in lower octave. In the avaroh with 140 Hz, we can see that the central part of the diagram is blank and the Middle octave notes have moved to upper corner and lower octave notes have moved near the right corner.

4.5.3 Tonic Selection

We have observed in the tristimulus analysis, singer's voice quality is best in the aaroh with 120 Hz tonic. But with this tonic he is not able to produce the lowest note of 80 Hz(Note 'm).

In the avaroh Singers voice quality is best with tonic 124.5 Hz. Lowest note with this tonic was 83.49 Hz. From the above discussions it is clear that singers voice range is 83.49 Hz to 346 Hz and his voice quality is best with the tonic greater than 124.5 Hz and less than 130 Hz.

Key #	Key color	Frequency (Hz)	Notation Used
1	White 1	240	C
2	Black 1	254	C # (D b)
3	White 2	269	D
4	Black 2	285	D # (E b)
5	White 3	302	E
6	White 4	320	F
7	Black 3	338.5	F # (G b)
8	White 5	358.5	G
9	Black 4	380	G # (A b)
10	White 6	402	A
11	Black 5	426	A # (B b)
12	White 7	451	B

Table 4.7: **Standard Middle C octave**

If we see the standard keys available on keyboard or harmonium, that are available with the pitches as shown in the table 1.3. These keys are reproduced below:

This is the middle octave; The corresponding lower octave pitches will be 120, 127,

134.5, 142.5, 151, 160, 169.25, 179.25, 190, 201, 213 and 225.5.

We can see in this range that the strong candidate for the tonic of this singer is 127 Hz.

We conclude that the suggested tonic for this singer is 127 Hz, which is commonly called as black 1 key.

4.6 Selection of Tonic: Rajendra Singh

Rajendra selected three tonics which were 123, 131, and 138 Hz. His notes in the aaroh were SA, RE, GA, ma, PA, DHA, NI, SA', RE', GA', ma', PA', and DHA' and in avaroh his notes were SA', 'NI, 'DHA, 'PA, 'ma, 'GA, 'RE. Note that all the notes are shuddha notes. We will analyze all these notes first to find out his voice range and then we will select tonic according to his voice quality using tristimulus diagrams.

Steady state parts of the notes were selected using pitch plots. Power spectrums of all these notes are shown in the figure. These power spectrums are analyzed for the various factors discussed in the previous sections.

4.6.1 Singer's Voice Range

In this subsection we will analyze notes in the aaroh and avaroh to find out singer's voice range. Analyzing aaroh we can get higher note and analyzing avaroh we can get lower limit of voice range similar to previous case. Various parameters are calculated for all the notes as shown in the tables.

Sr. No.	Note Name	Expected Pitch(Hz)	Observed Pitch(f_1)	RMSE	Harmonic Centroid($*f_1$)	Power in f_1 (%)	Avg. Power in Note
1	Sa	123.05	123.05	1.1153	2.5293	53.945	0.28113
2	Re	138.43	139.65	1.7747	3.3183	37.153	0.39822
3	Ga	153.81	155.76	1.5552	3.5339	23.235	1.1445
4	Ma	164.06	166.5	3.006	3.1433	25.682	1.1011
5	Pa	184.57	186.04	1.5323	2.8353	34.289	0.8729
6	Dha	207.64	208.98	2.9168	2.8929	26.334	0.91994
7	Ni	230.71	232.42	2.5824	2.5208	26.377	0.69522
8	Sa'	246.09	245.61	2.3944	2.4255	37.165	0.45626
9	Re'	276.86	278.81	3.6856	2.2166	27.953	0.82542
10	Ga'	307.62	312.01	6.1083	1.9318	35.127	1.0693
11	Ma'	328.13	331.54	5.2189	2.0754	25.199	1.8892
12	Pa'	369.14	371.09	3.0188	1.7246	36.087	1.4706
13	Dha'	415.28	423.34	5.7437	1.7316	54.359	0.62318

Table 4.8: **Aaroh 123 Raj**

4.6.1.1 Analysis of Aaroh

All the notes in the aaroh with three Sa are analyzed. Pitch plots, power spectrums are obtained as shown in the figures. From these power spectrums parameters shown in the table are obtained. As can be seen in the table and power spectrums, when the tonic is 123 Hz, highest note is 423.34 Hz which is reached comfortably with good harmonic structure. As can be observed in this table highest note with tonic 131 is 439.94 Hz. This note is attained quite comfortably with good harmonic structure. Highest note with tonic 138 Hz is 463.38 Hz whose power spectrum is shown in the figure 4.12. From the power spectrum it can be observed that the second harmonic is dominating. It is also clear that the fundamental and third harmonics are also considerably powerful. This shows that this note is reached comfortably with good harmonic structure. So we can consider highest note to be around this note.

Sr. No.	Note Name	Expected Pitch(Hz)	Observed Pitch(f_1)	RMSE	Harmonic Centroid($*f_1$)	Power in f_1 (%)	Avg. Power in Note
1	Sa	138.18	138.18	1.6128	3.3399	29.591	0.22196
2	Re	155.46	154.79	1.4666	3.8111	22.601	0.48582
3	Ga	172.73	171.39	1.77	3.3807	19.587	0.30066
4	Ma	184.24	182.13	2.4747	3.2739	26.204	0.2887
5	Pa	207.28	207.03	1.364	2.5888	39.35	0.1792
6	Dha	233.18	230.96	2.8227	2.6153	23.996	0.26465
7	Ni	259.09	258.3	1.47	3.1493	22.89	0.50014
8	Sa'	276.37	274.9	2.0162	2.3412	26.547	0.49696
9	Re'	310.91	305.18	6.4906	2.1062	24.085	1.0611
10	Ga'	345.46	340.33	5.1714	2.211	14.088	1.2497
11	Ma'	368.49	364.26	4.9674	1.8564	30.941	1.1065
12	Pa'	414.55	408.2	5.5444	1.6139	48.769	1.4442
13	Dha'	466.37	463.38	7.4647	1.8619	23.254	2.4747

Table 4.10: **Aaroh 138 Raj**

it is evident from the pitch plot singer is not able to sustain this note. Also it can be noted from the table that the power level of this note is too low. If we see the next higher notes, i.e. 'Ga and 'Ma, RMSE is considerable and Harmonic centroid is almost one. This indicates that only fundamental harmonic is present. According to Helmholtz, musical quality of such simple tones is dull.

Sr. No.	Note Name	Expected Pitch(Hz)	Observed Pitch(f_1)	RMSE	Harmonic Centroid($*f_1$)	Power in f_1 (%)	Avg. Power in Note
1	Sa	123.05	123.05	1.9807	2.8422	51.201	0.38232
2	'Ni	115.36	116.21	1.1504	2.1545	69.429	0.40115
3	'Dha	103.82	103.03	1.4257	1.2927	93.38	0.22817
4	'Pa	92.285	94.727	2.2409	1.07	92.997	0.10937
5	'Ma	82.031	90.332	233.77	1	100	0.016518

Table 4.11: **Avaroh 123 Raj**

So the singers lowest note is next higher note which is note 'Pa(104.98 Hz) in table 4.13, which has good harmonic centroid, low RMSE, and also it is powerful note. Also in table number 4.11 and 4.12, we can see at this pitch(103 and 108 respectively),these

Sr. No.	Note Name	Expected Pitch(Hz)	Observed Pitch(f_1)	RMSE	Harmonic Centroid($*f_1$)	Power in f_1 (%)	Avg. Power in Note
1	Sa	128.42	128.42	2.1931	2.6651	51.432	0.19923
2	'Ni	120.39	124.02	3.8109	2.618	53.044	0.26136
3	'Dha	108.35	110.35	1.2825	1.4137	91.174	0.22243
4	'Pa	96.313	98.633	2.1724	1.0274	97.256	0.21089
5	'Ma	85.612	91.309	7.0685	1.0761	92.389	0.089212

Table 4.12: **Avaroh 131 Raj**

Sr. No.	Note Name	Expected Pitch(Hz)	Observed Pitch(f_1)	RMSE	Harmonic Centroid($*f_1$)	Power in f_1 (%)	Avg. Power in Note
1	Sa	139.16	139.16	1.3373	3.5388	38.522	0.55704
2	'Ni	130.46	130.37	1.5197	4.3729	18.888	1.3537
3	'Dha	117.42	116.21	2.4402	2.622	61.65	0.3983
4	'Pa	104.37	104.98	1.443	1.3222	91.595	0.55608
5	'Ma	92.773	96.191	3.0184	1.0352	96.476	0.40697
6	'Ga	86.975	92.285	4.9122	1.0867	91.325	0.32072
7	'Re	78.278	84.961	251	1.093	90.699	0.055834

Table 4.13: **Avaroh 138 Raj**

parameters have comparable values. So we conclude from above discussion that the singers lowest note is around 105 Hz.

4.6.1.3 Singer's Voice Range

We can note from the previous two subsections that the voice range of this singer is:

Lowest Note is around 105 Hz and,

Highest Note is around 466 Hz.

4.6.2 Tristimulus Analysis

As in the previous section, using tristimulus diagram, three tonics are compared. Location of the notes in the tristimulus diagram specifies its harmonic structure and intern quality of note.

4.6.2.1 Tristimulus analysis of Aaroh

In the two tristimulus diagrams with tonics 123 Hz and 131 Hz, all the notes are in the central part of the diagram indicating good harmonic structure. Some of the higher octave notes are observed on central part of the y-axis. This indicates absence of higher harmonics but good balance of fundamental and mid band harmonics. As the tonic is increased now in the third tristimulus diagram we can observe that cluster of the notes is shifting upwards indicating decrease in amplitude of fundamental harmonic and dominance of mid band. This structure represents good harmonic structure in which all the three bands are reasonably powerful.

4.6.2.2 Tristimulus Analysis of Avaroh

We can observe in the tristimulus diagram of avaroh with tonic 123 Hz that only two notes i.e. Sa and 'Ni are away from the origin. All other notes are near about origin, indicating absence of higher harmonics other than fundamental. We have seen that the voice quality of this type of notes is dull. So tonic of the singer should be grater than this.

Now as tonic is increased, we can see in the tristimulus diagram of avaroh with Sa 131 Hz, that the cluster of the notes is shifted slightly away from origin, indicating increase in the number of harmonics present in the spectrum. Now, in the tristimulus diagram of avaroh with tonic 138 Hz, notes are shifted away from the origin into the central region, indicating good musical quality of these notes as compared to previous notes with only fundamental harmonic. Still some notes are very near to the origin.

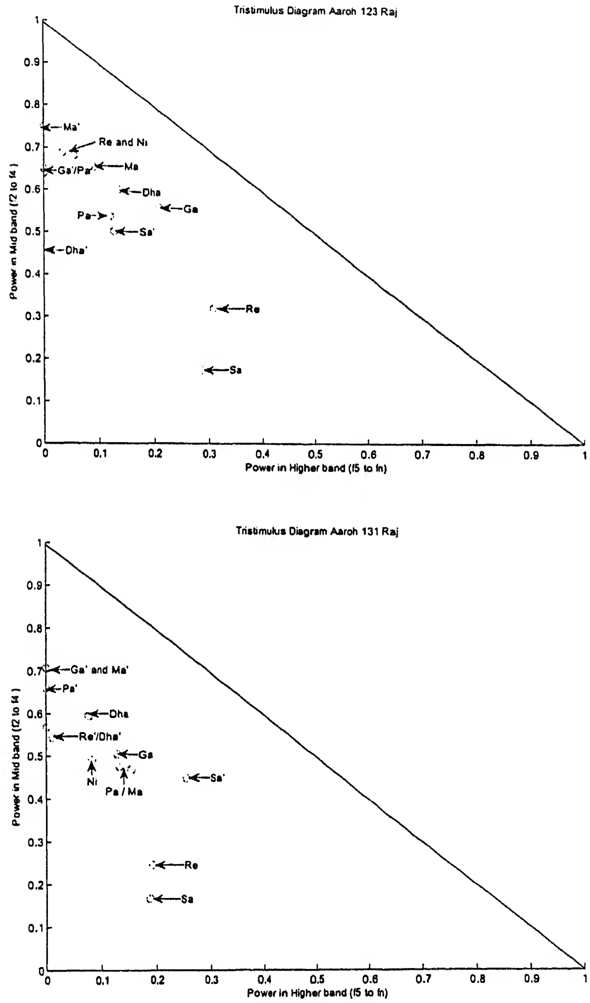


Figure 4.13: Spectral Variation: Aaroh tonic= 123 and 131 Hz, Singer: Rajendra Singh

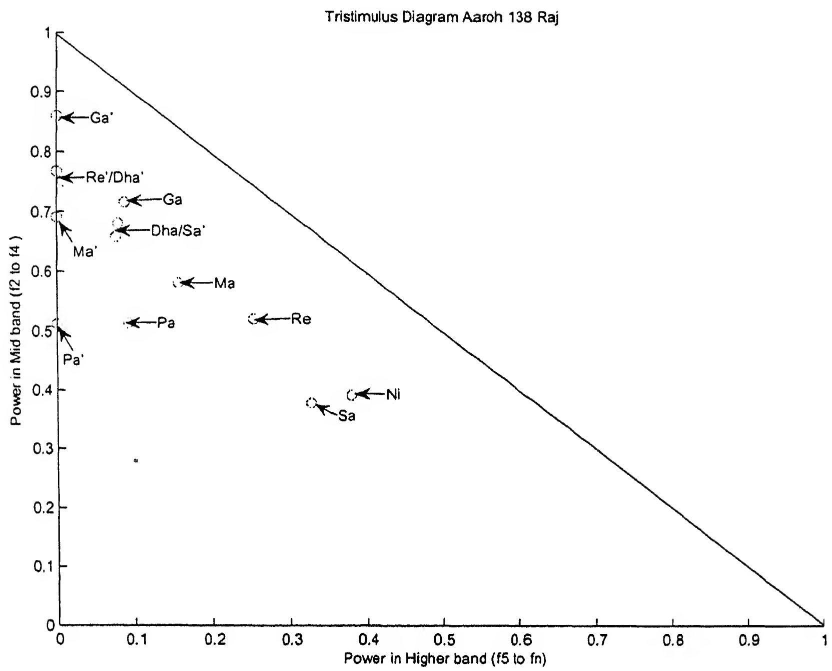
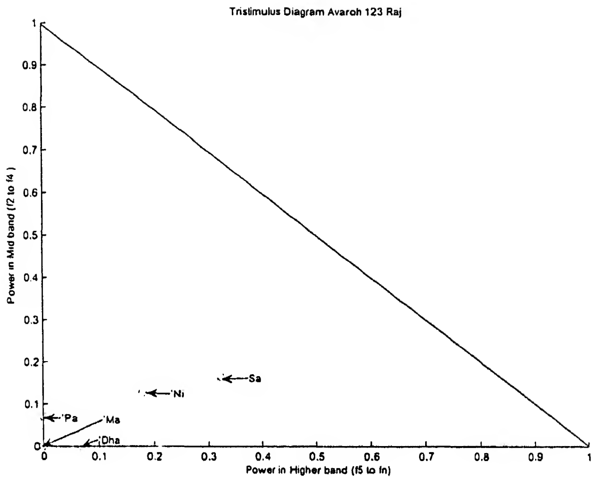


Figure 4.14: Spectral Variation: Aaroh tonic= 138 Hz, Singer: Rajendra Singh

4.6.3 Tonic Selection

From the tristimulus analysis it it clear that, in the aaroh singer has no problem with all the three tonics but in the avaroh with tonics 123 Hz and 131 Hz notes were too close to the origin, indicating only presence of fundamental and dull quality of notes. In the tristimulus diagram with tonic 138 Hz it was observed that the notes are shifted away from origin into the central part indicating increase in the number of harmonics. Still with this tonic some notes were near to the origin, indicating if further tonic is increased, these notes can shift away from this region and also it is clear that in the tristimulus diagram with tonic 138 Hz if, we further increase tonic by half note, in the



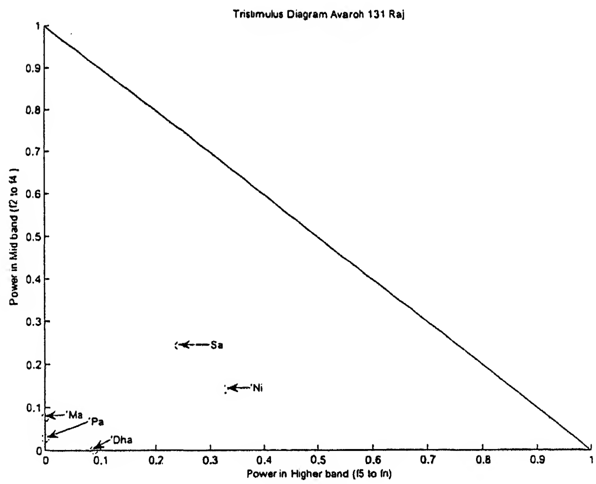


Figure 4.15: Tristimulus Diagram Avaroh,Tonic= 123, 131 Hz, Singer: Rajendra Singh

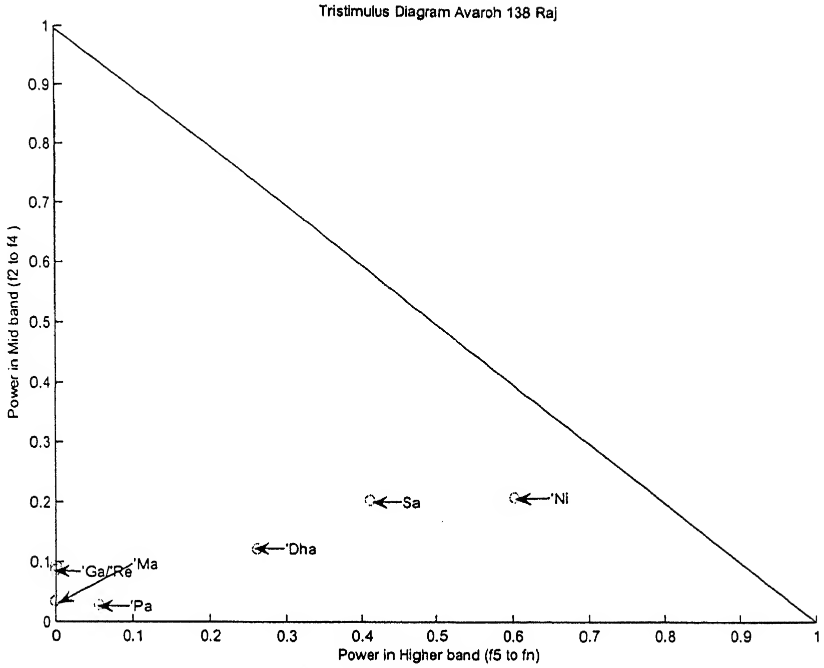


Figure 4.16: Spectral Variation: Avaroh tonic= 138 Hz, Singer: Rajendra Singh

tristimulus diagram notes will shift upwards slightly indicating further increase in the strengths of higher harmonics.

Singer's voice range, as we found in previous section is 105 to 466 Hz. We know from the tristimulus analysis best candidate for tonic is 138 Hz. Nearest keys available on keyboard or harmonium are 134.5 and 142.5. If we choose lower key, as observed in tristimulus diagrams of avaroh notes will be very close to origin, indicating presence of very strong fundamental and absence of other harmonics. On the other hand, if we choose 142.5 Hz notes in the avaroh will move towards the central part of the diagram. In the analysis of tristimulus diagram of avaroh with 138 Hz, it was observed

that still some notes were near the origin. In aaroh with 138 Hz, we found good harmonic structure. so if we choose tonic grater than 138 Hz, by half note i.e. 142.5 Hz, its tristimulus diagram will shift upwards which is required as seen in the previous sections.

If 142.5 Hz tonic is selected, singers notes in aaroh will be

Sr. No.	Note Name.	Ideal Pitch (Hz)
1	S	142.5
2	R	160.31
3	G	178.13
4	m	190
5	P	213.75
6	d	240.47
7	N	267.19
8	S'	285
9	R'	320.63
10	G'	356.25
11	m'	380
12	P'	427.5
13	D'	480.94

Table 4.14: **Aaroh with tonic 142.5**

and notes in the avaroh will be, From above table it is clear that singer is very

Sr. No.	Note Name.	Ideal Pitch (Hz)
1	S'	142.5
2	'N	133.59
3	'D	120.23
4	'P	106.88
5	'm	95
6	'G	89.063

Table 4.15: **Avaroh with tonic 142.5**

comfortable in lower pancham to upper pancham range. i.e. from 106.88 Hz to 427.5

Hz which is usually sufficient range for classical music and if required, singer can produce the other notes shown with less number of harmonics at lower and higher end.

So we conclude that recommended tonic for this singer is 142.5 Hz which is called as black 2 key.

Chapter 5

Conclusion & Future Work

The tonic of the singer mainly depends [9] on the two factors, one is his voice range and other, his voice quality over that range. In this thesis attempt is made to determine the voice range and tonic of the singer accurately by comparing voice quality(timbre) of singer with different tonics. Notes used for this comparison are represented by their steady state portions, That is the portion of the notes where pitch variation and spectral variation is less. So that we can represent that note by a single point in the tristimulus diagram.

The present study shows a method for tonic selection using signal processing techniques. Involvement of more number of musicologists, singers, trained listeners in the experiment can be very fruitful in the development of this method. So this is a task, which needs to be taken up immediately.

References

[1] Hermann L. F. Helmholtz, *On the Sensations of Tone as a Physiological Basis for the theory of Music* :Dover Publications, inc.,NY, 2nd edition, 1954.

[2] Chen-Gia Tsai, "Helmholt'z Nasality Revisited: Physics and Perception of Sounds with Predominance of Upper Odd-numbered Harmonics" in *Proceedings of the International Symposium on Musical Acoustics(ISMA 2004), Nara, Japan, April 2004.* , Department of Musicology, Humboldt University Berlin, Germany

[3] Arvindh Krishnaswamy, "Application of Pitch Tracking to South Indian Classical Music," in *Proc. IEEE IC ASSP 2003*, Hong Kong, China, Apr 2003.

[4] Arvindh Krishnaswamy, "Towards Modelling Computer Analysis and Synthesis of Indian Ragams," in *Proc. of FRSM 2004*, Chidambaram, India 2004.

[5] Arvindh Krishnaswamy, "Pitch Measurements Versus Perception of South Indian Classical Music," in *Proc. of the Stockholm Music Acoustics, 2003*, Stockholm, Sweden 2003.

[6] B. C. Deva, *Indian Music*:Indian council for cultural Realations,New Delhi, 1974, chap. 1–3, pp. 1–3877.

[7] A.K.Dutta, et.al, "Study of srutis in indian musical scales and relevance of schools in recent times using signal processing techniques," in *Proc. of FRSM 2003, Kanpur, India 2003.*

[8] Baburao Joshi, *Understanding Indian Music*: Asia Publising House, 1963, chap. 1–2, pp. 1–7.

[9] Alexander Wood, *The Physics of Music*: University Paperbacks,London,6th edition, 1964, chap. 1–5, pp. 1–77.

[10] Baburao Joshi,Antsher Lobo *Introducing Indian Music*: N.V alis Baburao Joshi,Kolhapur, edition 1965, chap. 1–3, pp. 1–18.

[11] Charles A. Culver, *Musical Acoustics*: MGH,inc,NY,4th edition, 1956, chap. 1,7–8, pp. 1–102.

[12] Russell L. Oertel and Pierce Cantrell, *LaTeX Example Manual.* Department of Electrical Engineering, Texas A&M University, College Station, Texas, September 1988.

[13] H.F. Pollard and E.V Jannson, "Tristimulus Method For The Specification of Musical Timbre,"*Acustica*, vol. v51, Issue N3, pp. 162–171, Aug. 1982.

[14] John fauvel, Raymond Flood, Robin Wilson,*Music And Mathematics*,Oxford University Press, NY, 2004.

[15] T. F. Quatieri, *Discrete-Time Speech Signal Processing*, Prentice Hall, 2002.

[16] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing: principles, algorithms and applications*, Macmillan, 1992.

[17] Alan V. Oppenheim, *Discrete-Time Signal Processing* , PHI, New Delhi, 2000.

[18] Alan V. Oppenheim,Ronald Schafer *Applications of Digital Signal Processing* , Prentice Hall, New Jersey, 1990.

[19] Wolfgang Hess, *Pitch and Voicing Determination* , University of Bonn, Bonn, Germany.

[20] S A K Durga, *Voice Culture*:Indian Musiocological Society,Mumbai, Vadodara, 1997, chap. 3-4, pp. 46–90.

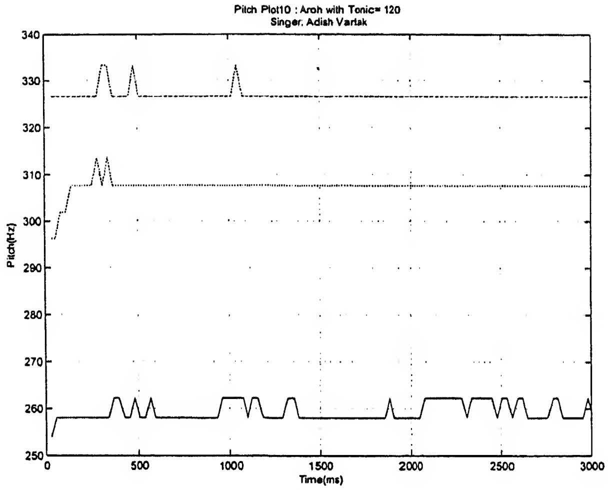
[21] http://www.soundofindia.com/

[22] http://www.khazanamusic.com/ramesh

[23] http://www.zainea.com/

Appendix A

Power Spectrums & Pitch Plots: Adish



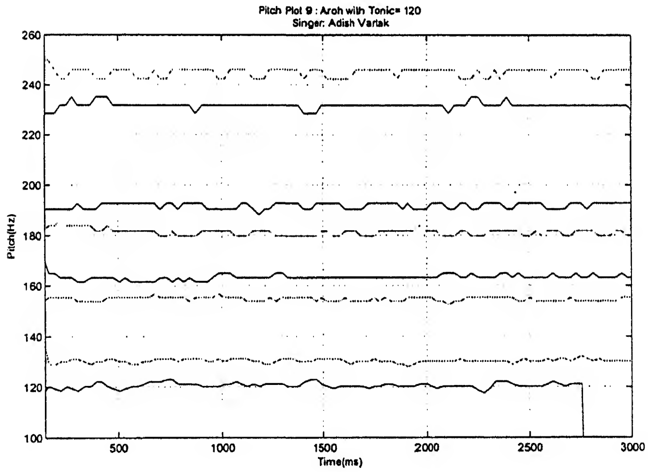


Figure A.1: Pitch Plot: Aaroh with Sa=120, Singer: Adish Vartak

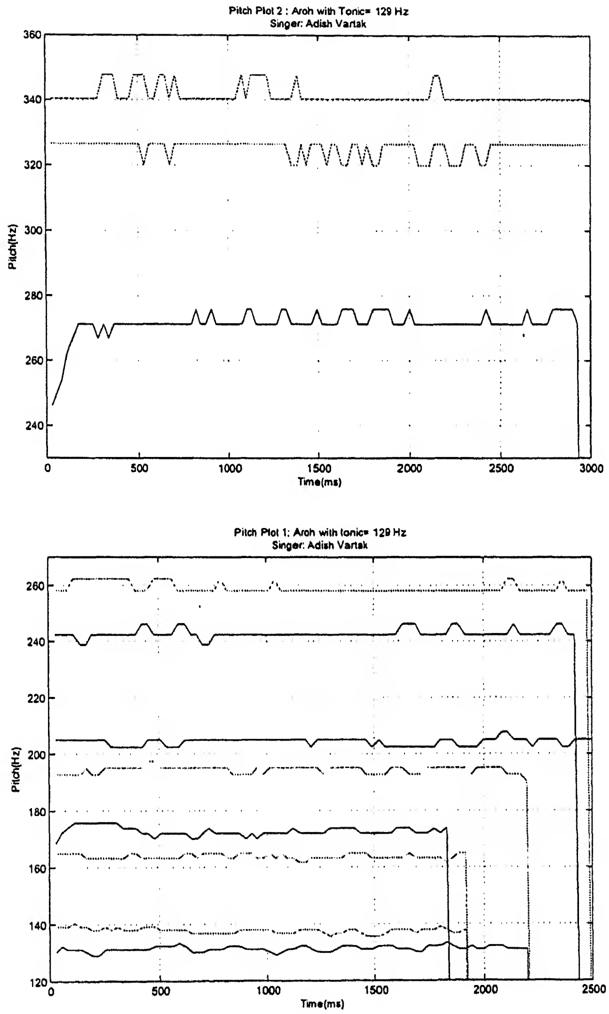


Figure A.2: Pitch Plot: Aaroh with Sa=129, Singer: Adish Vartak

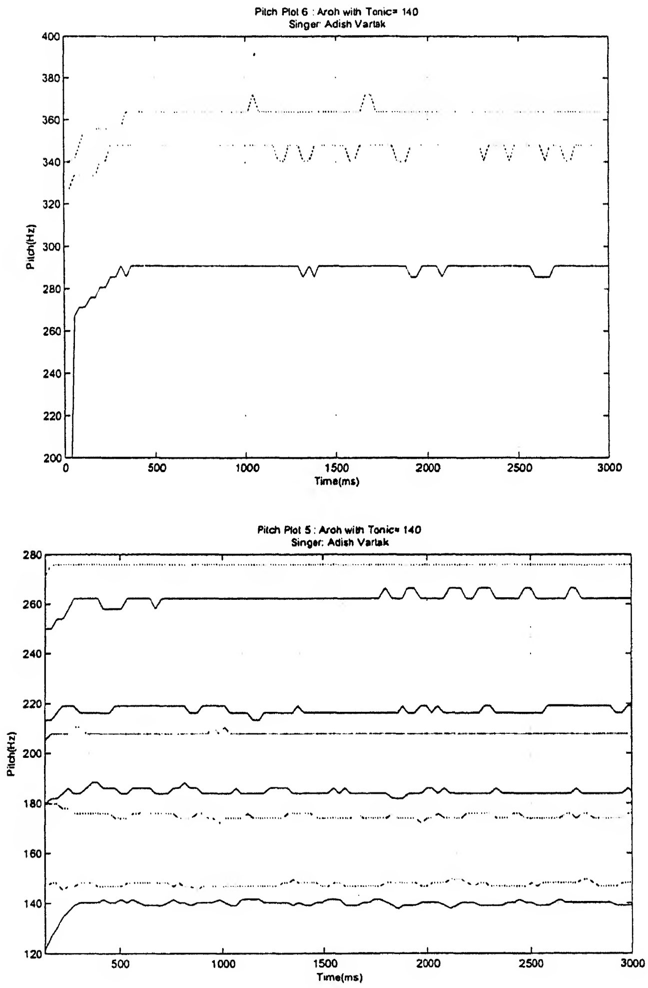


Figure A.3: Pitch Plot: Aaroh with Sa=140, Singer: Adish Vartak

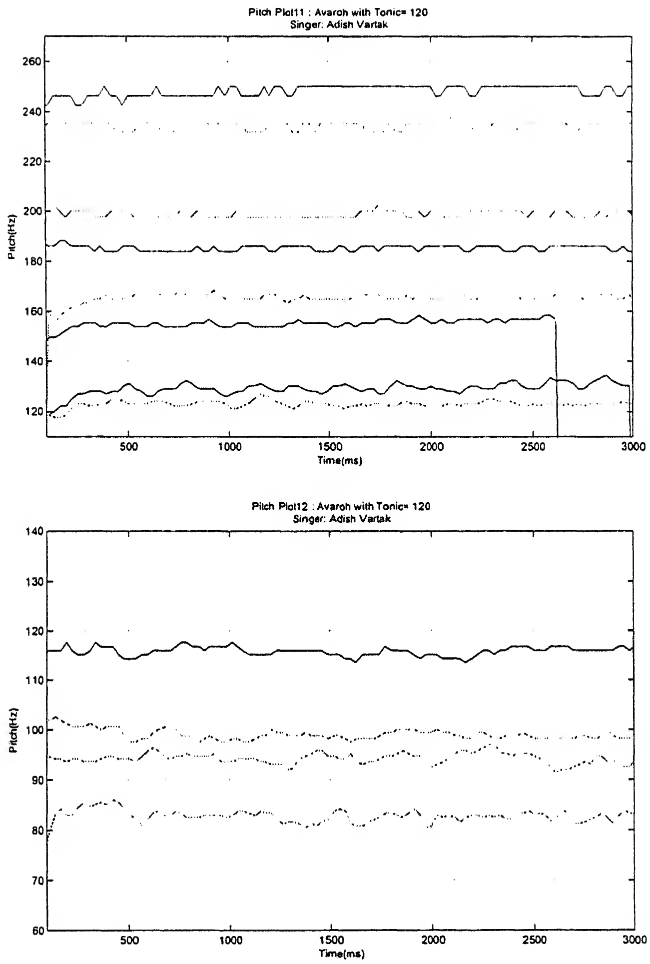


Figure A.4: Pitch Plot: Avaroh with Sa=120, Singer: Adish Vartak

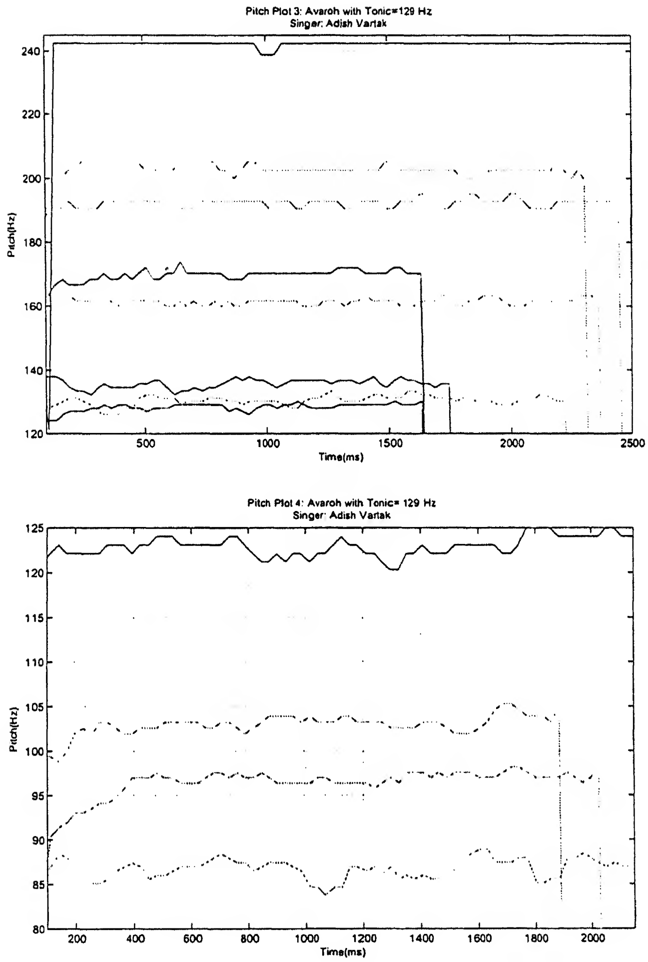
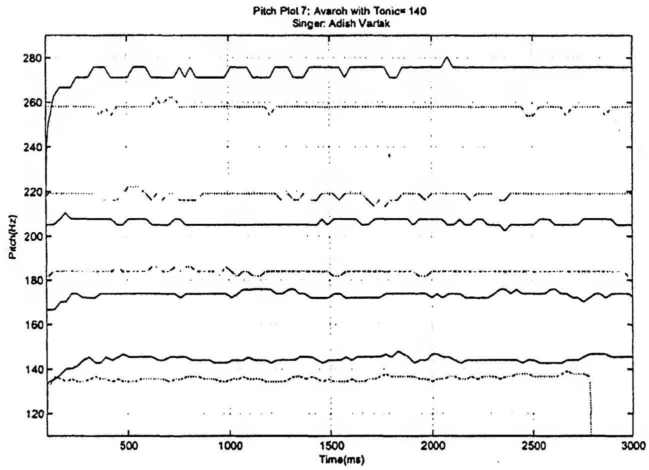


Figure A.5: Pitch Plot: Avaroh with Sa=129, Singer: Adish Vartak



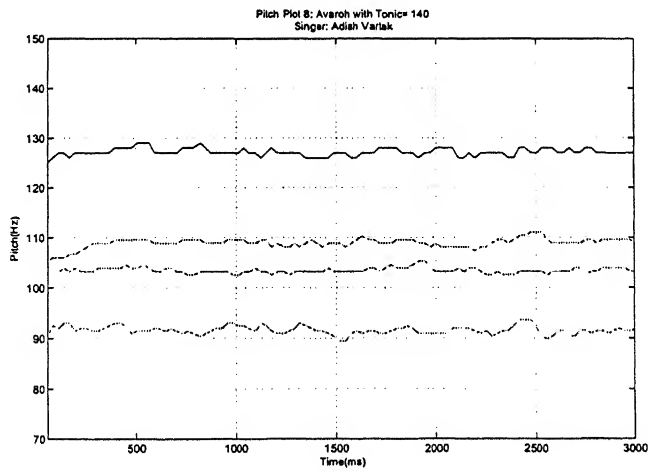


Figure A.6: Pitch Plot: Avaroh with Sa=140, Singer: Adish Vartak

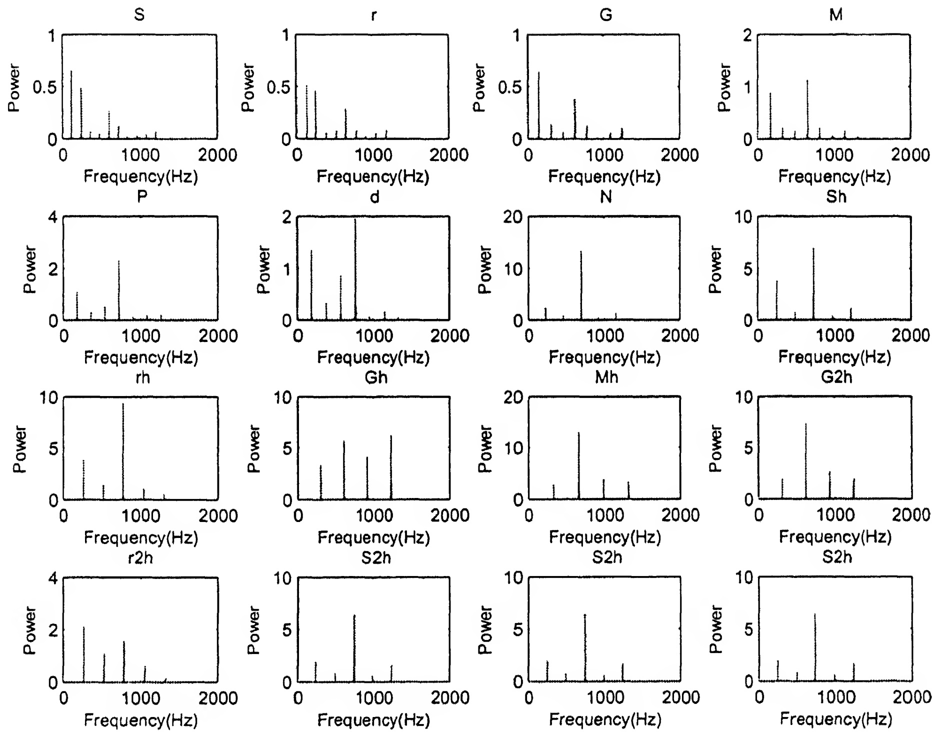


Figure A.7: Power Spectrum: Aaroh with Sa=120, Singer: Adish Vartak

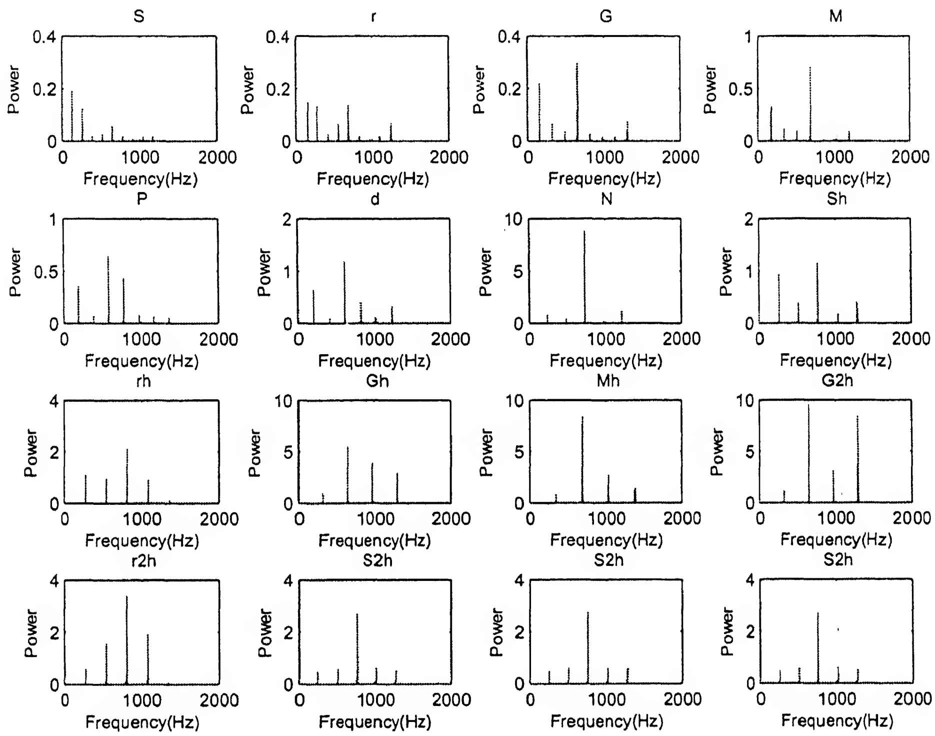


Figure A.8: Power Spectrum: Aaroh with Sa=129, Singer: Adish Vartak

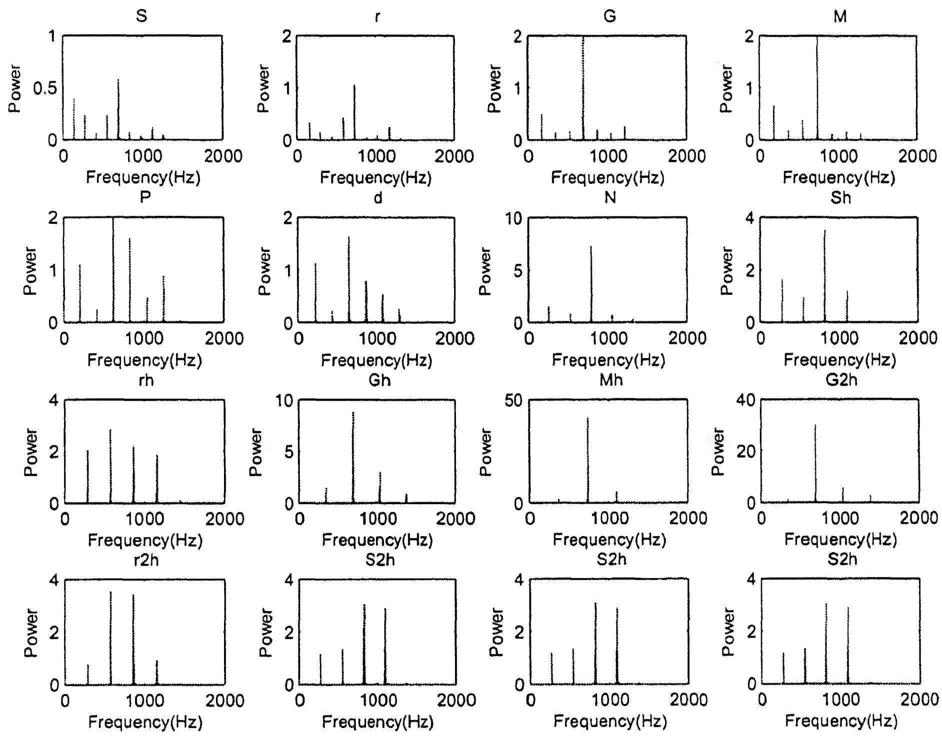


Figure A.9: Power Spectrum: Aaroh with Sa=140, Singer: Adish Vartak

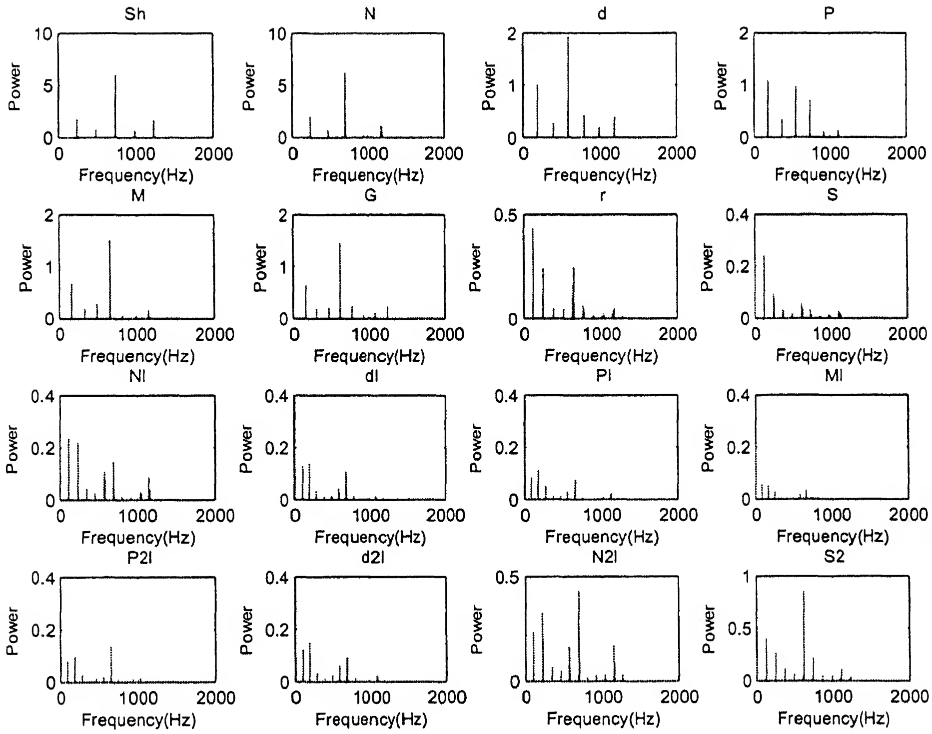


Figure A.10: Power Spectrum: Avaroh with Sa=120, Singer: Adish Vartak

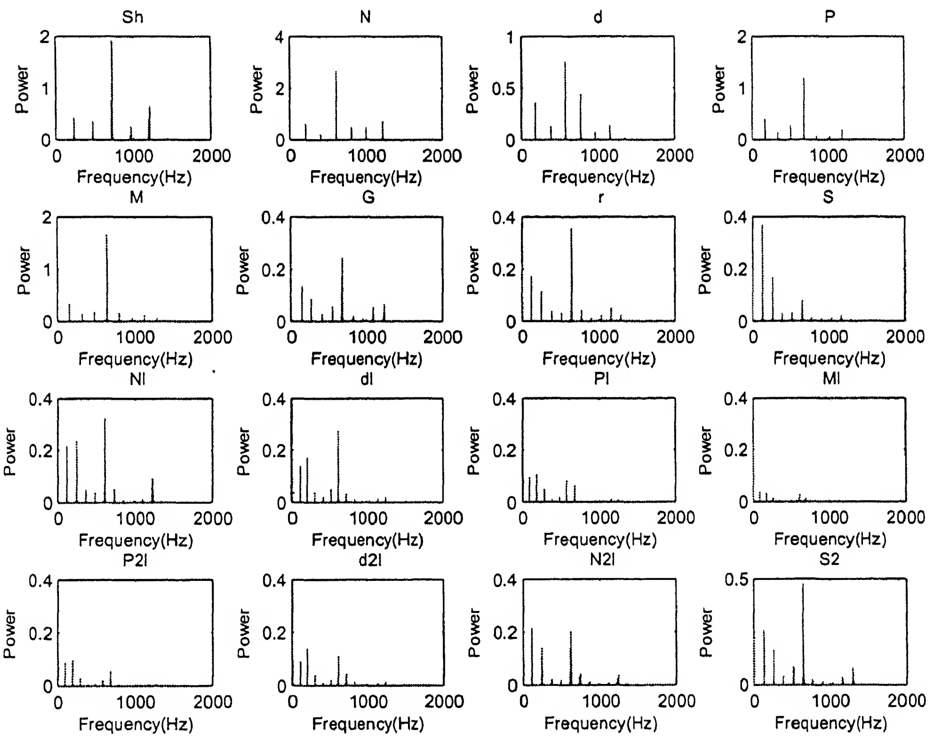


Figure A.11: Power Spectrum: Avaroh with Sa=129, Singer: Adish Vartak

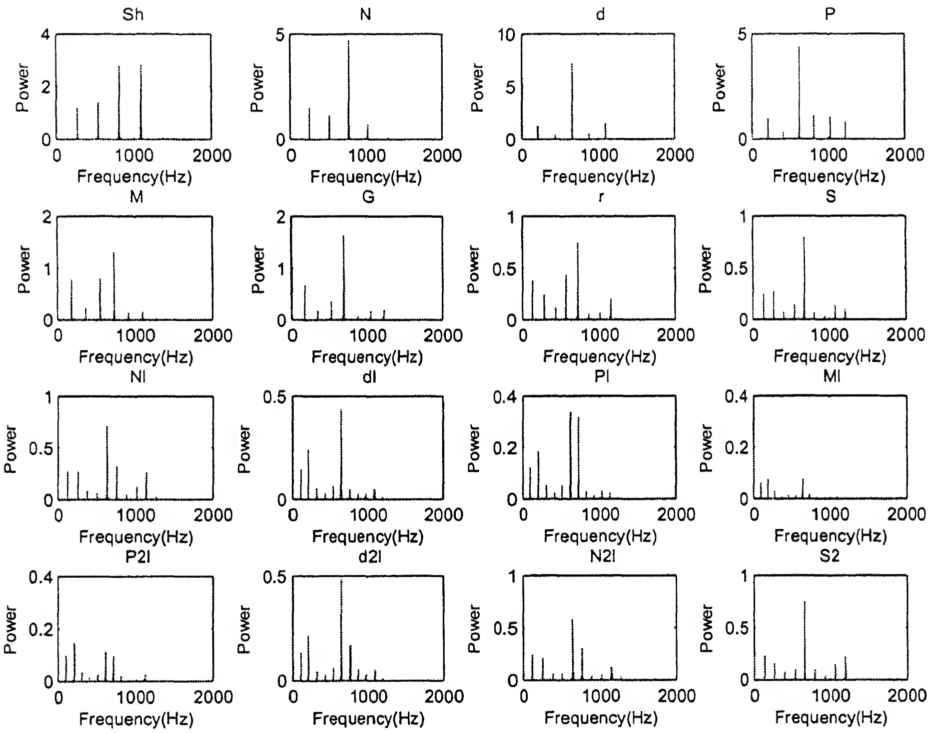


Figure A.12: Power Spectrum: Avaroh with Sa=140, Singer: Adish Vartak

Appendix B

Power Spectrums & Pitch Plots: Raj

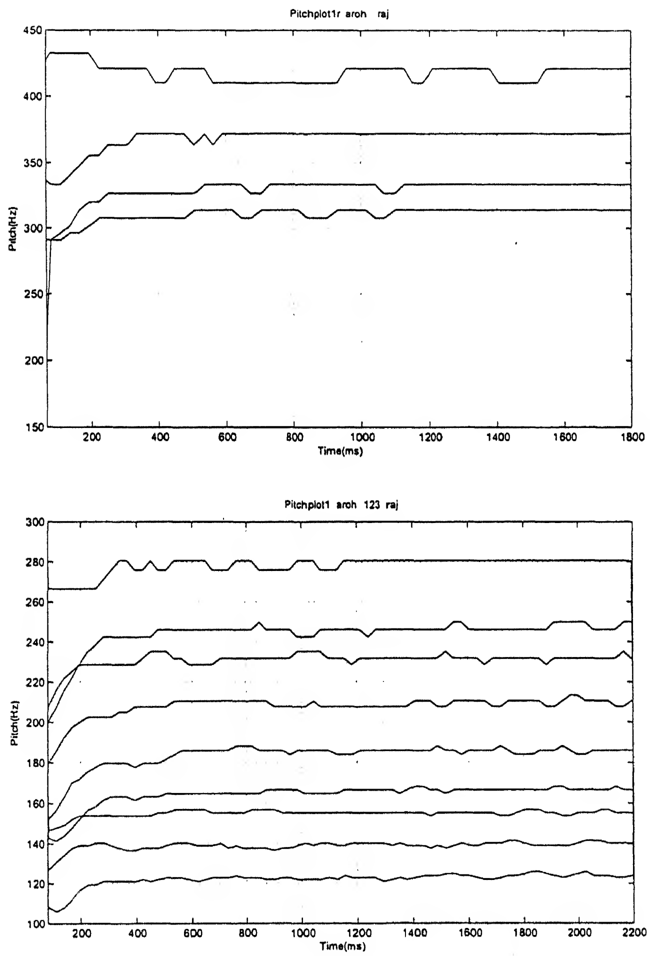


Figure B.1: Pitch Plot: Aaroh with Sa=123, Singer: Rajendra Singh

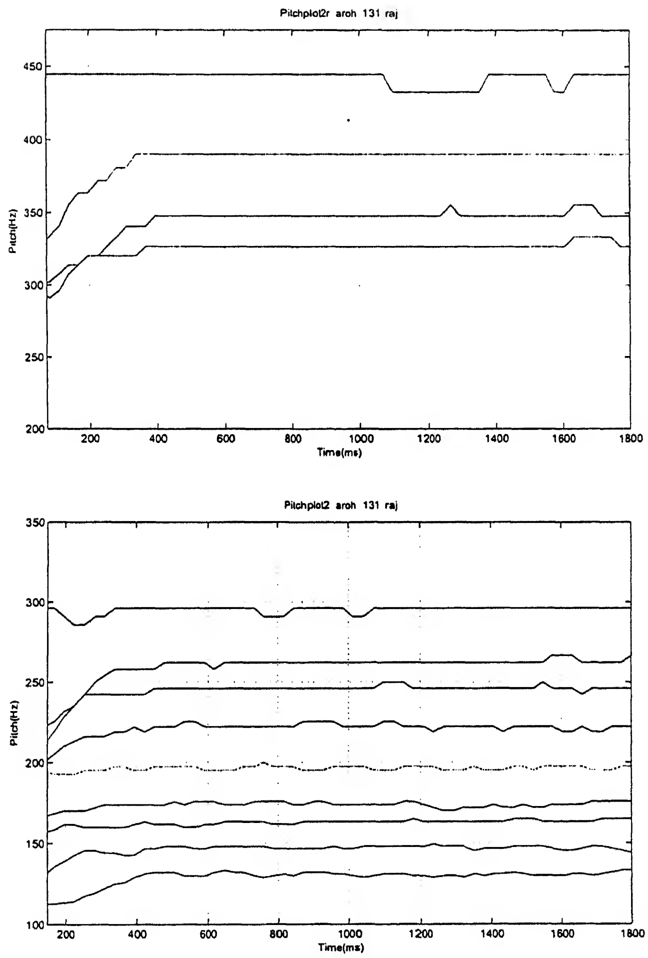
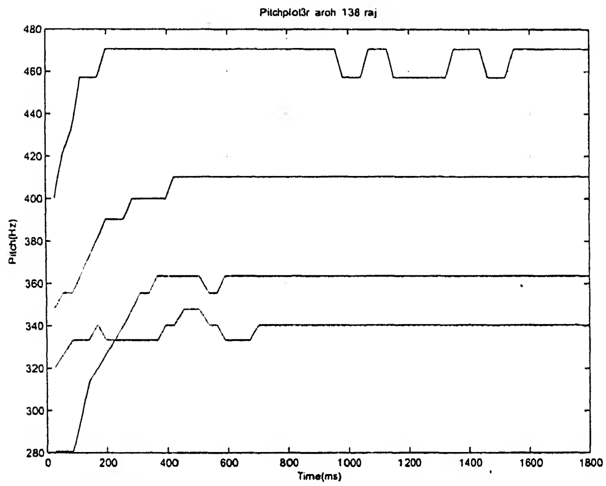


Figure B.2: Pitch Plot: Aaroh with Sa=131, Singer: Rajendra Singh



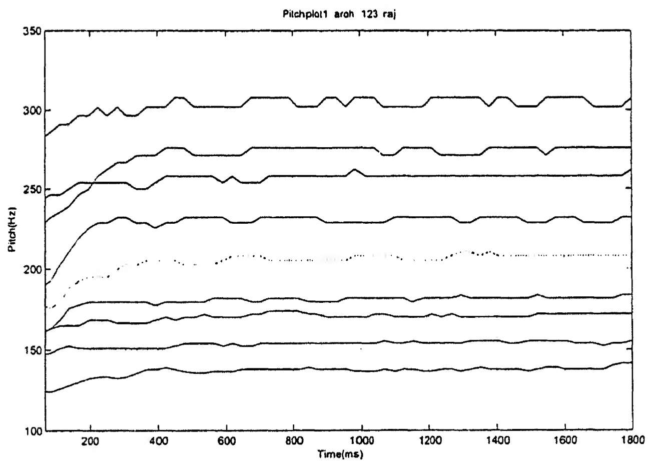


Figure B.3: Pitch Plot: Aaroh with Sa=138, Singer: Rajendra Singh

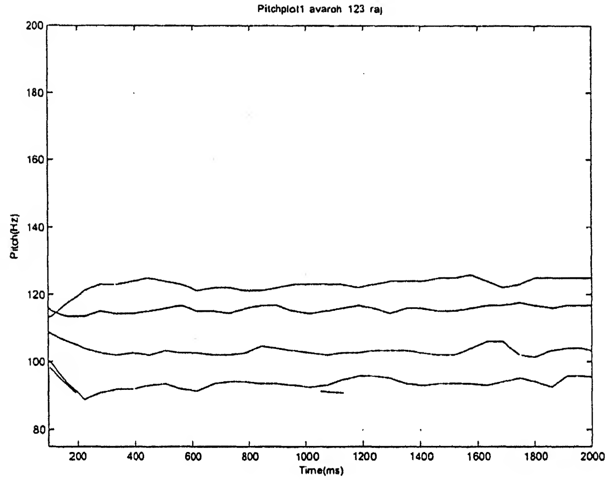


Figure B.4: Pitch Plot: Avaroh with Sa=123, Singer: Rajendra Singh

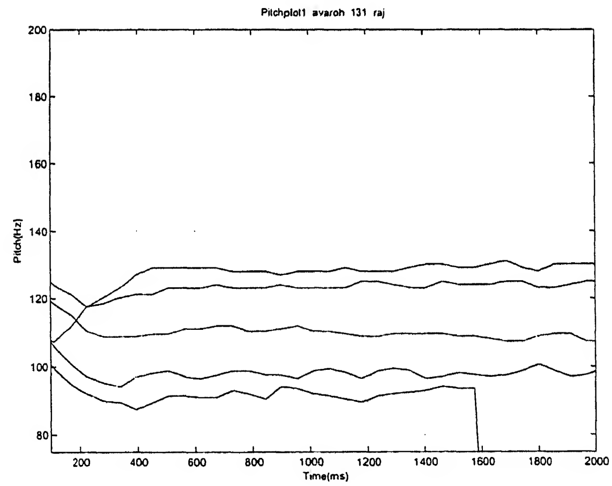


Figure B.5: Pitch Plot: Avaroh with Sa=131, Singer: Rajendra Singh

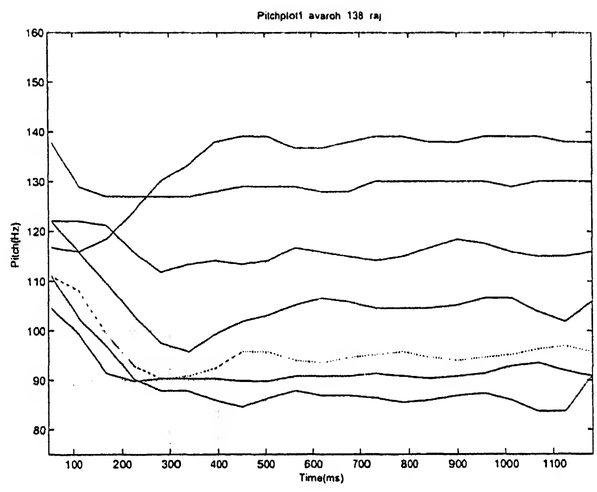


Figure B.6: Pitch Plot: Avaroh with Sa=138, Singer: Rajendra Singh

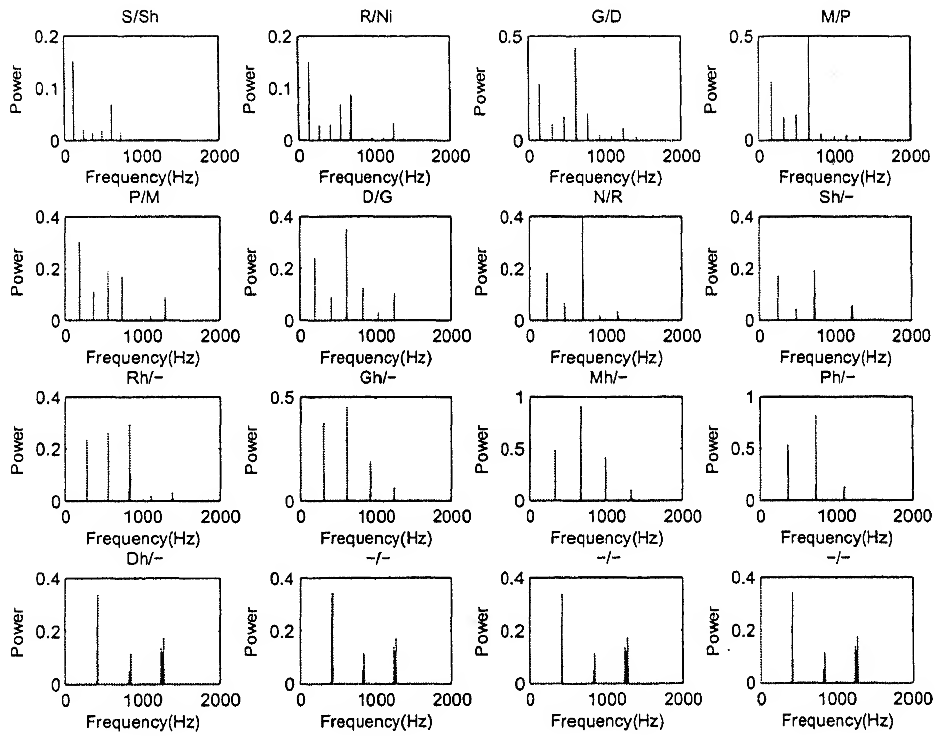


Figure B.7: Power Spectrum: Aaroh with Sa=123, Singer: Rajendra Singh

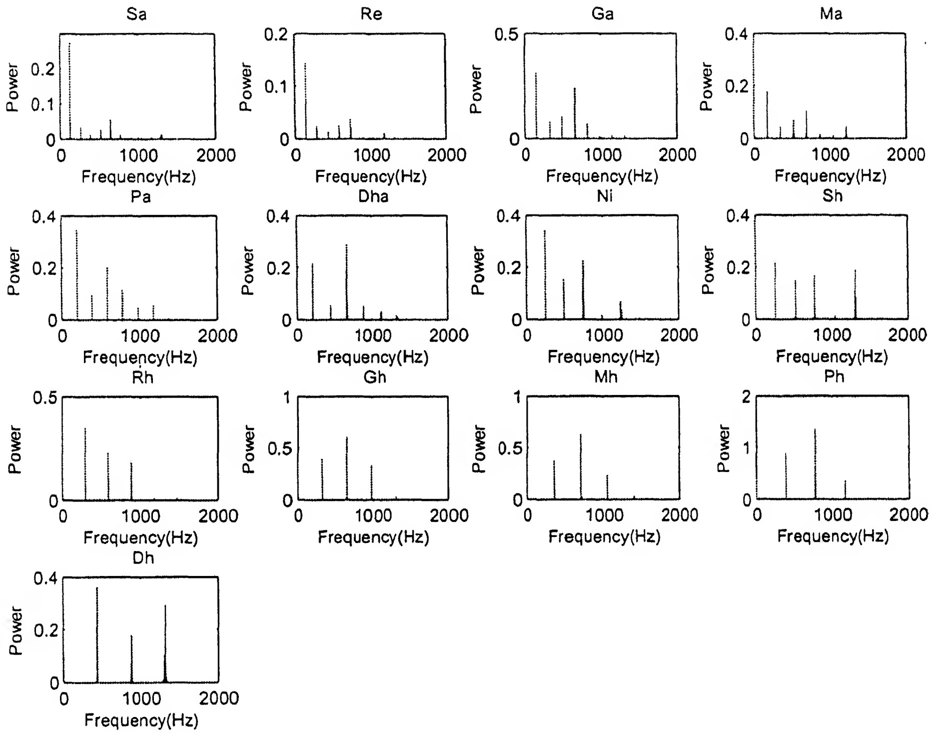


Figure B.8: Power Spectrum: Aaroh with Sa=131, Singer: Rajendra Singh

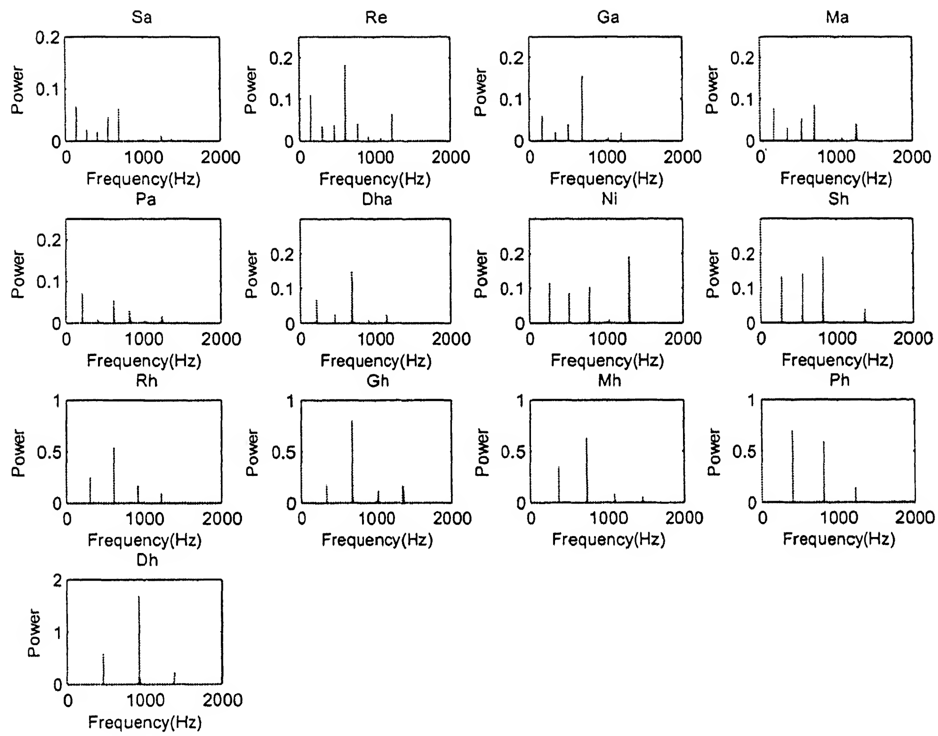


Figure B.9: Power Spectrum: Aaroh with Sa=138, Singer: Rajendra Singh

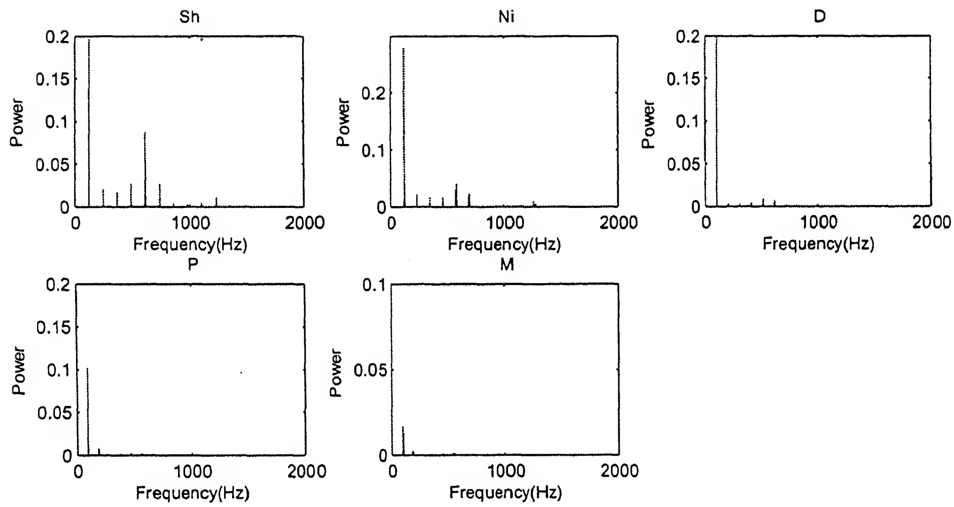


Figure B.10: Power Spectrum: Avaroh with Sa=123, Singer: Rajendra Singh

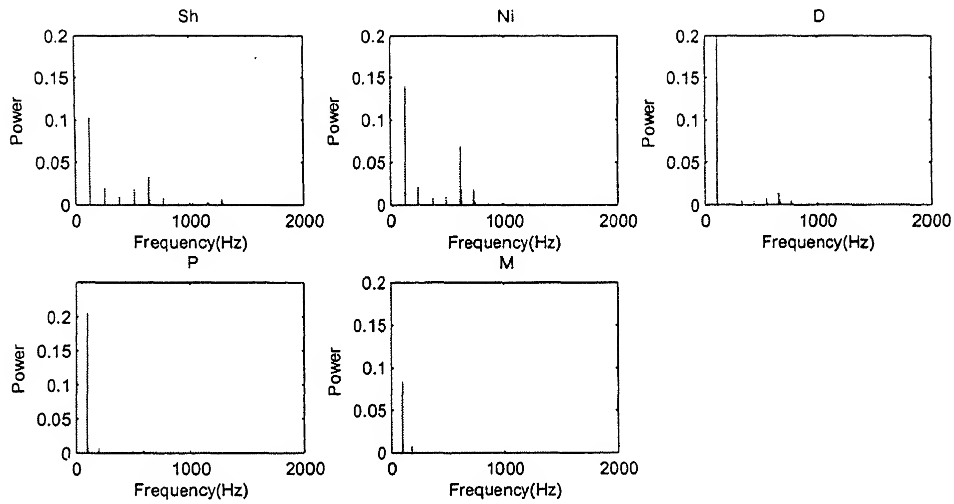


Figure B.11: Power Spectrum: Avaroh with Sa=131, Singer: Rajendra Singh

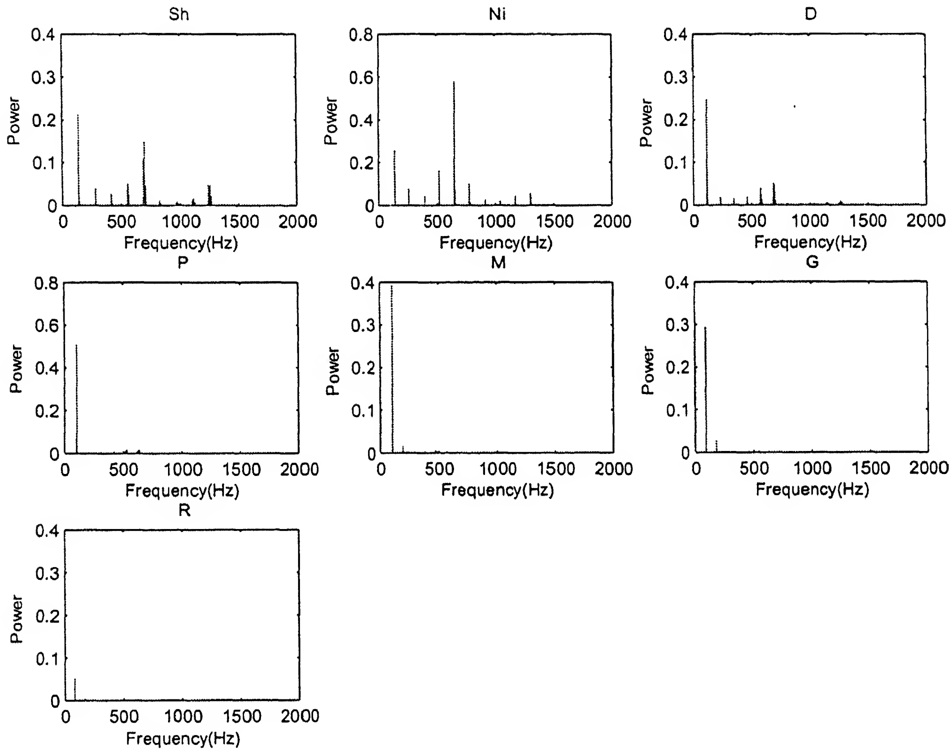


Figure B.12: Power Spectrum: Avaroh with Sa=138, Singer: Rajendra Singh